

OM of: US-09-303-518d-127 to: SPTREMBL\_19:\* out\_format: pfs:

Date: Jun 30, 2002 8:03 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet\_n2p.model -DEV=xlh  
-Q/cgn2.1/USPTO.spool/US09303518/runat\_28062002\_142714\_4338/app\_query.fasta\_1.23501  
-DB=SPTREMBL\_19 -PWT=fastan -SUFFIX=rspt -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=100 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USV=US09303518.ecgnt\_1\_1967 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-303-518d-127

Query length: 1344

Database: SPTREMBL\_19:\*

Database sequences: 562222

Database length: 172994929

Search time (sec): 883.700000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
sp_bacteria:Q44130	+ 1520.00	2343.45	7.8e-123	449	044130 actinobacillus pleuropne
sp_bacteria:Q9XG60	+ 636.00	970.69	2.2e-46	451	Q9XG60 porphyromonas gingivalis
sp_fungi:Q9C105	+ 139.50	189.08	0.0028	1236	09C105 schizosaccharomyces pom
sp_fungi:Q04051	+ 133.50	175.79	0.0106	1802	004051 saccharomyces cerevisia
sp_mammal:Q29071	+ 127.50	179.41	0.0228	528	Q29071 sus scrofa (pig). gastr
sp_invertebrate:P91365	- 124.50	159.56	0.0687	2232	P91365 caenorhabditis elegans
sp_human:Q99552	- 122.50	151.50	0.1207	3570	Q99552 homo sapiens (human). m
sp_human:Q9H9M1	- 122.00	168.86	0.0728	639	Q9H9M1 homo sapiens (human). n
sp_fungi:Q94317	+ 121.00	169.20	0.0834	534	Q94317 schizosaccharomyces pom
sp_bacteria:Q9WY86	+ 120.00	169.43	0.0959	451	Q9WY86 thermotoga maritima. ele
sp_human:Q14881	+ 118.00	162.93	0.1599	622	Q14881 homo sapiens (human). m
sp_archae:Q9HNR5	+ 117.00	163.81	0.1800	494	Q9HNR5 halobacterium sp. (stra
sp_invertebrate:Q9N5K0	- 116.50	153.65	0.2719	1203	Q9N5K0 caenorhabditis elegans
sp_human:Q9UQ36	+ 116.50	153.04	0.2775	1275	Q9UQ36 homo sapiens (human). m
sp_human:Q15038	+ 116.50	149.50	0.3123	1783	Q15038 homo sapiens (human). m
sp_human:Q60382	+ 116.50	149.46	0.3128	1791	Q60382 homo sapiens (human). m
sp_human:Q9UQ35	+ 116.50	144.93	0.3638	2752	Q9UQ35 homo sapiens (human). m
sp_invertebrate:Q9U0V2	- 116.00	146.48	0.3719	2207	Q9U0V2 leishmania major. pos
sp_human:Q96HAI	- 113.50	151.11	0.4605	984	Q96HAI homo sapiens (human). m
sp_invertebrate:Q9XK60	- 112.00	153.46	0.5310	631	Q9XK60 mus musculus (mouse). m
sp_plant:Q9SU0V	+ 110.00	147.13	0.8807	857	Q9SU0V arabidopsis thaliana (mc
sp_invertebrate:Q9BKV7	- 110.00	142.54	1.03	1325	Q9BKV7 leishmania major. pos
sp_mammal:Q29070	+ 109.00	156.06	0.7575	317	Q29070 sus scrofa (pig). gastr
sp_bacteria:P94909	+ 108.00	149.11	1.11	529	P94909 microbacterium ammoni
sp_mammal:Q9N1P0	+ 108.00	144.52	1.29	818	Q9N1P0 bos taurus (bovine). su
sp_virus:Q93781	+ 108.00	143.92	1.32	866	Q93781 equine herpesvirus 1. me
sp_invertebrate:Q73749	- 108.00	140.31	1.48	1219	Q73749 xenopus laevis (afri
sp_human:Q9GZV6	- 107.50	139.82	1.62	1187	Q9GZV6 homo sapiens (human). m
sp_human:Q14395	- 107.00	148.05	1.33	505	Q14395 homo sapiens (human). m
sp_invertebrate:Q9V515	- 107.00	143.94	1.52	746	Q9V515 drosophila melanogast
sp_human:Q9Y2N2	- 107.00	138.13	1.85	1294	Q9Y2N2 homo sapiens (human). m
sp_invertebrate:Q9VXN5	- 107.00	131.59	2.30	2406	Q9VXN5 drosophila melanogast
sp_bacteria:Q9T3F4	+ 106.50	151.08	1.29	352	Q9T3F4 pseudomonas aeruginosa
sp_bacteria:Q05089	+ 106.50	146.47	1.51	545	Q05089 nocardioides simplex (ar
sp_virus:Q93782	+ 106.50	141.57	1.78	867	Q93782 equine herpesvirus 1. me
sp_invertebrate:Q76602	- 106.50	137.51	2.03	1275	Q76602 caenorhabditis elegans
sp_bacteria:p73758	+ 105.50	144.37	1.87	574	p73758 synecocystis sp. (stra
sp_human:Q9UQ39	+ 105.50	138.99	2.24	956	Q9UQ39 homo sapiens (human). m
sp_human:Q9UQ40	+ 105.50	136.07	2.47	1262	Q9UQ40 homo sapiens (human). m
sp_human:Q9UHA8	+ 105.50	129.76	3.05	2296	Q9UHA8 homo sapiens (human). s

seq_name: sp_bacteria:Q44130	seq_documentation_block:	ID	AC	Q44130	PRELIMINARY:	PRT:	449 AA.
sp_bacteria:Q99UW0	+	105.00	146.52	1.88			
sp_mammal:Q18758	-	105.00	110.48	6.25			
sp_invertebrate:Q9NK94	+	104.50	142.69	2.30			
sp_human:Q43419	-	104.50	142.54	2.31			
sp_human:Q9H195	-	104.50	138.06	2.68			
sp_human:Q9P237	-	104.00	139.12	2.79			
sp_invertebrate:Q9QY35	-	104.00	128.99	3.91			
sp_invertebrate:Q9N8G4	-	103.50	133.07	3.67			
sp_bacteria:Q9RYM2	+	103.00	139.10	3.23			
sp_invertebrate:Q9QW42	-	103.00	127.10	4.82			
sp_virus:Q91TW8	+	102.50	143.03	3.05			
sp_bacteria:Q52747	+	102.50	137.92	3.62			
sp_human:Q9UKW9	-	102.50	131.79	4.44			
sp_invertebrate:Q9N8G4	-	102.50	125.10	5.55			
sp_invertebrate:Q9N3Y8	-	102.50	124.88	5.59			
sp_plant:Q9M715	+	102.00	146.73	2.90			
sp_bacteria:Q9RYP2	+	102.00	143.39	3.24			
sp_bacteria:Q44512	+	102.00	139.68	3.67			
sp_human:Q14879	-	102.00	139.52	3.69			
sp_human:Q9UF83	-	102.00	138.82	3.78			
sp_human:Q9NYE4	-	102.00	136.60	4.07			
sp_virus:Q93307	-	102.00	136.12	4.13			
sp_human:Q15329	-	101.50	141.43	3.73			
sp_bacteria:Q92C76	+	101.50	141.10	3.77			
sp_human:Q12955	-	101.50	116.75	8.50			
sp_bacteria:Q936S5	+	101.00	138.69	4.40			
sp_invertebrate:Q70511	-	101.00	121.37	7.84			
sp_fungi:Q74851	-	100.50	129.46	6.44			
sp_bacteria:Q9KA24	+	100.50	138.75	5.08			
sp_bacteria:Q98K15	+	100.00	136.97	5.39			
sp_plant:Q9C548	-	100.00	131.41	6.50			
sp_invertebrate:Q9V957	-	99.50	122.75	9.33			
sp_invertebrate:Q22579	-	99.50	122.75	9.33			
sp_human:Q43421	-	99.00	137.05	6.23			
sp_invertebrate:Q9QX5	+	99.00	120.67	10.77			
sp_invertebrate:Q70670	+	99.00	120.18	10.95			
sp_invertebrate:Q9QX47	+	99.00	119.18	11.32			
sp_bacteria:Q92TC9	+	98.50	137.15	6.69			
sp_bacteria:Q92T02	+	98.50	137.12	6.69			
sp_invertebrate:Q95TR0	-	98.50	123.01	10.72			
sp_invertebrate:Q969D4	-	98.50	115.46	13.80			
sp_bacteria:Q98CW8	+	98.00	133.93	8.02			
sp_human:Q14120	+	98.00	125.14	10.75			
sp_human:Q9U9V0	+	98.00	120.29	12.64			
sp_invertebrate:Q9PV22	-	98.00	119.37	13.03			
sp_fungi:Q96D11	-	98.00	118.97	13.21			
sp_plant:Q9SH38	-	97.50	130.88	9.55			
sp_invertebrate:Q9VI63	-	97.50	130.52	9.67			
sp_invertebrate:Q9NHX4	-	97.50	123.14	12.37			
sp_invertebrate:Q9BM29	-	97.50	122.73	12.54			
sp_plant:Q23054	-	97.00	134.52	8.99			
sp_bacteria:Q93RW3	+	97.00	134.82	9.02			
sp_bacteria:Q74024	+	97.00	131.95	9.92			
sp_mammal:Q9M2L1	-	97.00	131.56	10.05			
sp_bacteria:Q9RW96	-	97.00	130.11	10.55			
sp_human:Q15052	-	97.00	123.00	13.38			
sp_fungi:Q96P11	-	97.00	118.06	15.78			
sp_human:Q9H4A3	-	97.00	116.17	16.81			
sp_archae:Q58289	-	96.50	141.94	7.65			

seq\_name: sp\_bacteria:Q44130  
seq\_documentation\_block:  
ID Q44130  
AC Q44130  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE 48 KDA OUTER MEMBRANE PROTEIN.  
GN AOPA.  
OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Actinobacillus.  
 RX NCBI\_TaxID=715;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=ISU158;  
 RX MEDLINE=96110919; PubMed=8557378;  
 RA Cruz W.T., Nedialkov Y.A., Thacker B.J., Mulks M.H.;  
 "Molecular characterization of a common 48-kilodalton outer membrane  
 protein of Actinobacillus pleuropneumoniae.";  
 RT Infect. Immun. 64:83-90(1996).  
 RL EMBL: U24492; AAC43631.1; -;  
 DR EMBL: U24492; AAC43631.1; -;  
 SQ SEQUENCE 449 AA; 48604 MW; 62CBDD17A8435418 CRC64;

alignment\_scores:  
 Quality: 1520.00 Length: 450  
 Ratio: 3.979 Gaps: 3  
 Percent Similarity: 84.889 Percent Identity: 66.667

alignment\_block:  
 US-09-303-518D-127 x Q44130 ..

Align seg 1/1 to: Q44130 from: 1 to: 449

```

1 ATCATTAATAATCAAAAGAGTCTAAACCTGCCATCGCGGACACCGGA 50
|||||
1 MetilethrIleuLysGlyLeuAspLeuProIleAlaGlyThrProAl 17
|||||
51 GCAAGTCATTATGACGGCCCGCTCATATTACCGAAGTCGCGTTGCTGGCG 100
|||||
17 aGlnValIleHisAsnGlyAsnThrValAsnGluValAlaMetLeuGlyG 34
|||||
101 AAGAAATATGCCGTATGCCGCCCTNGATCAAAAGTCAAGGAAGCGCATGCC 150
|||||
34 luGluTyrrValGlyMetA-rProSerMetLysValArgGluGlyAspVal 50
|||||
151 GTCAAAAAGGCCCAAGTGGTGTGTTGAAGACAAAGATCCGGGGTGGT 200
|||||
51 ValLysGlyGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67
|||||
201 GTTTACGGCCCGNGTTTACGGCAAAATCGCGCCATCCATCGCGGGA 250
|||||
67 lPheThrAlaProAlaSerGlyThrValValThrIleAsnArgGlyGluL 84
|||||
251 AGCGGTFACCTTCAGTCGGTGGTTCGATTCGCGTTCGAAGCAACACGA 300
|||||
84 ysArgValLeuGlnSerValValIleLysValGluGlyAspGluGlnIle 100
|||||
301 GAGTTCGAACGCTACGCGCCCGGAGCGTTGGCAAACTTAAGCGCGGANG 350
|||||
101 ThrPheThrArgTyrrGluAlaAlaGlnLeuAlaSerLeuSerAlaGluG 117
|||||
351 ANTNNNGNCAATCTGATCCAAATCCGGTTTGTGGACTGCGCTGCGPANCC 400
|||||
117 nValLysGlnAsnLeuIleGluSerGlyLeuThrPheThrAlaPheArgTh 134
|||||
401 GTCCGTTTCAGCAAAATCCCTCGCGTGGATGCCGAGCGGTCGCCATCTTC 450
|||||
134 rgProPheSerLysValProAlaLeuAspAlaIleProSerSerIlePhe 150
|||||
451 GTCATCGCGATGGACACCAATTCGCTNGCGGCGACACCCCTGTGGTTGTAT 500
|||||
151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProGluValValIle 167
|||||
501 CAAAGAACCGCNGANGATTTTCAGACGANGTNTGCTGGTATTGAGCGGTT 550
|||||
167 uLysGluTyrrGluThrAspPheLysaspGlyLeuThrValLeuThrArgL 184
|||||
551 TG.....ACGAGCGTAAATCCATGTGTGTAAGCAGCTGCCGCGAC 594
|||||
184 euPheAsnGlyGlnLysProValTyrrLeuCysLysAspAlaAspSerAsn 200
|||||
595 GTGCGG...TCTGAAATGCTGCAACATCGAATCAACATCAATTCGGCGG 641

```

## RP SEQUENCE FROM N.A.

RC STRAIN=W50;  
 RA Ross B.C., Barr I., Patterson M., Agius C., Rothe L., Margetts M.,  
 RA Hocking D., Webb E.;  
 RT "Porphyromonas gingivalis polypeptides and nucleic acids."  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF14076; AAD33930.1;  
 SQ SEQUENCE 451 AA; 49825 MW; 3E4CC9B66FE5AD74 CRC64;

## alignment\_scores:

Quality: 636.00 Length: 452  
 Ratio: 2.208 Gaps: 7  
 Percent Similarity: 63.717 Percent Identity: 34.071

## alignment\_block:

US-09-303-518D-127 x Q9X6S0

Align seg 1/1 to: Q9X6S0 from: 1 to: 451

1 ATGATTAATAAATCAAAAAGGCTTAACCTGCCATCGCCGCGGACACGC... 48  
 :::::||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 4 VallileysThrlsLysGlyLeuAlaLeuMetHisLysLysProLe 20  
 49 .GAGCAAGTCAATTATGACGCGGCCCGCTTACCGAAGTCCGCTGCTG 97  
 :::::||||| ||||| ||||| ||||| ||||| ||||| |||||  
 20 uProGluMetLeuAlaGluProAlaGlnSerProThrTyrAlaValp 37  
 98 CGAAGATATGCGCGTATGCGCCCTNGATGAAGTCAAGGAGCGCAT 147  
 :::::||||| ||||| ||||| ||||| ||||| ||||| |||||  
 37 roAspPheGluGlyValLeuProLysValThrAlaArgProGlyAsp 53  
 148 GCCGTCAAAAAGGCCAAGTCTGTTTGAAGACAAAGATCGGCGCT 197  
 :::::||||| ||||| ||||| ||||| ||||| ||||| |||||  
 54 LysValargAlaGlySerAlaLeuMetHisLysLysAlaTyrProGluMe 70  
 198 GGTGTTTACCGCCGCTTTCAGCAAAATCGCCCATCCATCGCGCG 247  
 :::::||||| ||||| ||||| ||||| ||||| ||||| |||||  
 70 tLysPheThrSerProValSerGlyGluValLeuAlaValAsnArgGlyA 87  
 248 AAAAGCGGTACTTCAGTCTGCTGATGCGGTTGAAGCAAGCAGCAA 297  
 ||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||  
 87 laLysArgLysValLeuSerileGluValLysProAspGlyLeuAsnGlu 103  
 298 ATCGAG...TTCGAACGCTACGCGCCGCGAAGCGTTGGCAAACTTAAGCG 344  
 ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 104 TyrGluSerPheProValGlyAspProSerAla.....LeuSerAl 117  
 345 CGANGAANTNNGNCAATCTGATCCAAATCCGTTTGTGGACTCGCGTGC 394  
 ::::: ||||| ||||| ||||| ||||| ||||| |||||  
 117 aGluGlnIleLysGluLeuLeuSerSerGlyMetTyrPheIleL 134  
 395 GTANCCGTCGCTCAGCAAAATCCCTCCGTCGATCGCGAGCGGTGCGC 444  
 ::::: ||||| ||||| ||||| ||||| ||||| |||||  
 134 ysGlnArgProTyrAspIleValAlaThrProAspIleAlaProArgAsp 150  
 445 ATCTTCGTCATGCGATGGACACAAATCCGCTNGCGCAGACCCCTGTGGT 494  
 ||||| ::::: ||||| ||||| ||||| ||||| |||||  
 151 IleTyrIleThrAlaAsnPheThrAlaProLeuAlaProAspPheAspPh 167  
 495 TGTGATCAAGAGCCGCGCANGATTTACAGCANGTNTGCTGGTATTGA 544  
 ::::: ||||| ||||| ||||| ||||| ||||| |||||  
 167 eIleValArgGlyGluGluArgAlaLeuGlnThrAlaIleAspAlaLeuA 184  
 545 GCGTTTACCGCGGTAAATCCATGTGTGTAAGCAGCTGCGCGCAGAC 594  
 ::::: ||||| ||||| ||||| ||||| ||||| |||||  
 184 laLysLeuThrThrGlyLysValTyrValGlyLeuLysProGlySerSer 200  
 595 GTGCGGCTCTGAAATGCTGCCAATCGCAACATCGAAACACATGATTCGCGCGCC 644  
 ::::: ||||| ||||| ||||| ||||| ||||| |||||  
 201 LeuGlyLeuHisAsnAlaGluIleValGluValHis.....GlyPr 214  
 645 GCATCGCGCGGCTTGTAGTGGCAGCACATTCATTTTATGAGCGCGTGC 694

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 214 oHisProAlaGlyAsnValGlyValLeuIleAsnHisThrLysProIleA 231  
 695 GTCAACAAACAAACCGTTTGGACCATCAATATCAAGATGTAATGCGCATC 744  
 ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 231 snArgGlyGluThrValTyrThrLeuLysAlaThrAspLeuIleValle 247  
 745 GGACGTTTGTTCACAGCGCTCTGAACACGCGCGCTGATTCCTTT 794  
 ||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||  
 248 GlyArgPheLeuLeuThrGlyLysAlaAspPheThrArgMetIleAlaMe 264  
 795 GGGTGCTTCTCAAGTCAACAAACCGCTCTCTGCTACCGTTTGGGTG 844  
 ||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||  
 264 tThrGlySerAspAlaAlaHisGlyTyrValArgIleMetProGlyC 281  
 845 CGAAGTATCGCAATTTACTGGCGGCAATTTGTT.....GACGCGAC 888  
 ::::: ||||| ||||| ||||| ||||| ||||| |||||  
 281 ysAsnValPheAlaSerPheProGlyArgLeuThrIleLysGluSerHis 297  
 889 AACCGCTGATTTCCGTTCCGTTTGAACGGCGGATTCACACAGGCGC 938  
 ::::: ||||| ||||| ||||| ||||| ||||| |||||  
 298 GluArgValIleAspGlyAsnValLeuThrGlyLysLeuCysGlu 314  
 939 GCACGATTTTGGGACGCTACCAATCAGATTTCCGTTATTCGAAGAAG 988  
 ::::: ||||| ||||| ||||| ||||| ||||| |||||  
 314 sGluProPheLeuSerAlaArgCysAspGlnIleThrValIleProGluG 331  
 989 GCCGACG...AAAGAGCTGTTCCGCTGGTTCGCCGCGCAGCGGACAAA 1035  
 ||||| ::::: ||||| ||||| ||||| ||||| |||||  
 331 LysAspValAspGluLeuPheGlyTyrAlaAlaProArgLeuAspGln 347  
 1036 TACTCCATCAGCGTACGACCTCGCCCATTTCCCTG...AAAAACAAC 1082  
 ||||| ::::: ||||| ||||| ||||| ||||| |||||  
 348 TyrSerMetSerArgAlaTyrPheSerTyrLeuGlnGlyLysAsnLysG 364  
 1083 CTTCAAGTTACGACGCGCTCAACGGTGGCGACCGCCCATGTCGCGA 1132  
 ::::: ||||| ||||| ||||| ||||| ||||| |||||  
 364 uTyrValLeuAspAlaArgIleLysGlyGlyGluArgAlaMetIleMetS 381  
 1133 TTGGTACTTACGACGCGCTAATCGCCTAGACATCTGCTACCTCGCTT 1182  
 ::::: ||||| ||||| ||||| ||||| ||||| |||||  
 381 erAsnGluTyrAspArgValPheProMetAspIleTyrProGluTyrLeu 397  
 1183 TTGCGCGATTTAATCGTCGCGCATACCGACGCGCGCAGCATTCGGTGT 1232  
 ||||| ::::: ||||| ||||| ||||| ||||| |||||  
 398 LeuLysAlaIleAlaPheAspLysMetGluAspLeuGlyI 414  
 1233 CTTGGAATTTGGACGAAGACCTCGCTTGTGCGAGCTTCGCTCGCCG 1282  
 ||||| ||||| ||||| ||||| ||||| |||||  
 414 eTyrGluValAlaProGluAspPheAlaThrCysGluPheValAspThrS 431  
 1283 GCAATACGAATANGCGCGCTGTTGCGTAAGTGTCTGGAACACCTTGA 1332  
 ::::: ||||| ||||| ||||| ||||| ||||| |||||  
 431 erLysIleGluLeuGlnArgIleValArgGluGlyLeuAspMetLeuTyr 447  
 1333 AAGGAA 1338  
 |||||  
 448 LysGlu 449

seq\_name: sp\_fungi:Q9C105

seq\_documentation\_block:

ID Q9C105 PRELIMINARY; PRT; 1236 AA.  
 AC Q9C105;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE PUTATIVE GLUCAMYLASE I (ALPHA-L-4-GLUCAN GLUCOSIDASE), EXTRACELLULAR  
 DE STARCH-DEGRADING ENZYME, BY SIMILARITY TO S. CEREVISIAE STAL, CONTAINS  
 DE CHITINASE FAMILY SIGNATURE.  
 GN SPAPBIE7.04C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;









```
468 GGTGTCATCGCATGACGAAGATGCGAAGCGCTCGGCATCGACGGCAG 419
:   :   :   :   :   :   :   :   :   :   :   :   :   :
918 eSerThrThrSerGluGluMetThrSerClnGlySerThrGlnThrPro. 934
:   :   :   :   :   :   :   :   :   :   :   :   :   :
418 GGATTTGCTGAACGACGAGCGGNTACGAGCGGAGTCCACAAACCGGATTGG 369
:   :   :   :   :   :   :   :   :   :   :   :   :   :
935 .....GlySerThrGlySerThrValThrGlnProSerThr 946
:   :   :   :   :   :   :   :   :   :   :   :   :   :
368 ATCAGATTCCNNCANNATTCN...TCGCGCGCTTAAGTTTGCCACGCGTTC 322
:   :   :   :   :   :   :   :   :   :   :   :   :   :
947 ValSerAspSerThrSerSerGlySerThrValThrValGlySerThrGln 963
:   :   :   :   :   :   :   :   :   :   :   :   :   :
321 GGGCGCGTAGCTAGCTGCAATTCGATTTCGCTTCGCTTCAACGGCAATCA 272
:   :   :   :   :   :   :   :   :   :   :   :   :   :
963 uGly.....SerSerSerProIleProSerThrSerGlnA 975
:   :   :   :   :   :   :   :   :   :   :   :   :   :
271 CGACCGACTGAGTAGCTAGCGCTTTTCGCCCGATGGATGGCGCGGATTGG 222
:   :   :   :   :   :   :   :   :   :   :   :   :   :
975 snThrAsnProSerThrSer...SerGlySerSerMetSerThrGlnThr 990
:   :   :   :   :   :   :   :   :   :   :   :   :   :
221 CCTGAACACGCGC.....GGGTAAACACACCGCCCGGATN 187
:   :   :   :   :   :   :   :   :   :   :   :   :   :
991 ProGlnSerSerClnSerThrSerProValGluSerSerThrSerGlyAl 1007
:   :   :   :   :   :   :   :   :   :   :   :   :   :
186 CTTTTCCTCTCAACAGCAGCACTGGCGCTTTTGTGACGGCA.....T 146
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1007 aThrSerSerSerGlySerProGlyThrThrLeuThrSerIleSerProS 1024
:   :   :   :   :   :   :   :   :   :   :   :   :   :
145 CGCCTTCCTTGACTTTCATCNCAGGGCGGATACCGGCATATCTTCGCGCA 96
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1024 erProSerProSerThrIleGlySerSerClnGlySerThrSerPro 1040
:   :   :   :   :   :   :   :   :   :   :   :   :   :
95 AGCAACGCGACT..... 84
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1041 ValValSerThrIleSerGlnGlySerThrGluThrProGlySerThrGln 1057
:   :   :   :   :   :   :   :   :   :   :   :   :   :
83 .TCGGTAATGACGGGCCGCTCAATAATGACTTGCTCCGCTGCGCCGGCA 35
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1057 ySerThrValThrLysProSer.....ThrValSerGlySerAlaSerS 1072
:   :   :   :   :   :   :   :   :   :   :   :   :   :
34 TGGGC 30
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1072 erGly 1073
:   :   :   :   :   :   :   :   :   :   :   :   :   :
seq_name: sp_human:Q99552
seq_documentation_block:
ID Q99552 PRELIMINARY; PRT; 3570 AA.
AC Q99552;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MUCIN MUC5B (FRAGMENT).
GN MUC5B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=97166151; PubMed=9013550;
RA Desseyn J.L., Guyonnet-Duperat V., Porchet N., Aubert J.P., Laine A.;
RT "Human mucin gene MUC5B, the 10.7 kb large central exon encodes
RT various alternate subdomains resulting in a super-repeat. Structural
RT evidence for a 11p15.5 gene family.";
RL J. Biol. Chem. 272:3168-3178(1997).
DR EMBL; Z72496; CAA96577.1; -.
FR NON_TER 1
FT NON_TER 3570
SQ SEQUENCE 3570 AA; 361019 MW; DE04E4D72759312 CRC64;
```

```
alignment_scores:
  Quality: 122.50      Length: 482
  Ratio: 0.544        Gaps: 21
  Percent Similarity: 46.680      Percent Identity: 22.822

alignment_block:
US-09-303-518D-127/rev x Q99552 ..
Align seg 1/1 to: Q99552 from: 1 to: 3570

1337 TCCTTCTCAANGTTTCCAGCACCTTACGCAACAGCGGGCCNTATTCGTA 1288
:   :   :   :   :   :   :   :   :   :   :   :   :   :
2882 SerThrAlaThrProSerSerThrProGlyThrAlaProProProLysVa 2898
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1287 TTTG.....CCGCGGACAGACGAAGCTG.....CACAAAGCGAGGT 1253
:   :   :   :   :   :   :   :   :   :   :   :   :   :
2898 lLeuThrSerProAlaThrThrProThrAlaThrSerSerLysAlaThrS 2915
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1252 CTCTTCGTCTCC..... 1233
:   :   :   :   :   :   :   :   :   :   :   :   :   :
2915 erSerSerSerProArgThrAlaThrThrLeuProValLeuThrSerThr 2931
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1232 CAACCCAAATGCTCGCGCTGTCGCTATCGCGACGATTAAATCGCGCAA 1183
:   :   :   :   :   :   :   :   :   :   :   :   :   :
2932 AlaThrLysSerThrAlaThrSerValThrProIleProSerSerThrLe 2948
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1182 AAGCAGGCTAGCGAGGATGCTACGGGCATTACGGCTCGTAAGTACCAA 1133
:   :   :   :   :   :   :   :   :   :   :   :   :   :
2948 uGlyThrThrGlyThrLeuProGluGlnThrThr.....ThrProV 2962
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1132 TCGGCACCATGGCGCGCTCGCCACCGTTGACGGCTGCTGCTGAACCTGAAG 1083
:   :   :   :   :   :   :   :   :   :   :   :   :   :
2962 alAlaThrMetSerThrIleHisProSerSerThr..... 2973
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1082 AGTTTGTTCAGAAATGCGCGAGGTCGTACGCGTGATGGAGTATTT 1033
:   :   :   :   :   :   :   :   :   :   :   :   :   :
2974 .....ProGluThrHisThrSerThrValLe 2983
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1032 GTCCGGCTGCGCGCAACCCAGCGCAAGCTCTTTCTCGCGGCTTCCT 983
:   :   :   :   :   :   :   :   :   :   :   :   :   :
2983 uThrLysAlaThrThrArgAlaThrSerSerSerThrSerThrProSerS 3000
:   :   :   :   :   :   :   :   :   :   :   :   :   :
982 CG.....ATAACGGAAATCTGATTGTGGTAGCGT 954
:   :   :   :   :   :   :   :   :   :   :   :   :   :
3000 erThrProGlyThrThrTripleLeuThrGluLeu..... 3011
:   :   :   :   :   :   :   :   :   :   :   :   :   :
953 CCCAAATAATCGTCGCGCTTGTGTAATCGCGCTTCAATACCGAAC 904
:   :   :   :   :   :   :   :   :   :   :   :   :   :
3012 .....ThrThrAlaAlaThrThrThrAlaGlyThrGlyPr 3023
:   :   :   :   :   :   :   :   :   :   :   :   :   :
903 GGAAATCACGGGTGTCT..... 885
:   :   :   :   :   :   :   :   :   :   :   :   :   :
3023 oThrAlaThrProSerSerThrProGlyThrThrTripleLeuThrGluL 3040
:   :   :   :   :   :   :   :   :   :   :   :   :   :
884 .....GCCTCAACCAATTCGCCGCGAGTAATT 858
:   :   :   :   :   :   :   :   :   :   :   :   :   :
3040 euThrThrThrAlaThrThrThrAlaSerThrGlySerThrAlaThrLeu 3056
:   :   :   :   :   :   :   :   :   :   :   :   :   :
857 TCGCATACTTTTCGCCACCCCAACCGGTACGCAAGAGCGCTGTTTGTGAC 808
:   :   :   :   :   :   :   :   :   :   :   :   :   :
3057 SerSerThr.....ProGlyThrThr.....TrpIleLeuTh 3067
:   :   :   :   :   :   :   :   :   :   :   :   :   :
807 TTGAGAACCCCAACGAATCAGCGCTCGGCTTCAGACGGCCTTTG 758
:   :   :   :   :   :   :   :   :   :   :   :   :   :
3067 r...GluProSerThrThrAlaThrValThrAlaProProGlySerThrA 3083
:   :   :   :   :   :   :   :   :   :   :   :   :   :
757 CAAACAAACGCTCCGATGCAATTTACATCTTGATAATTGATGGTCCAAACG 708
:   :   :   :   :   :   :   :   :   :   :   :   :   :
3083 laThrAlaSerSerThrGlnAlaThrAlaGlyThrProHisValSerThr 3099
:   :   :   :   :   :   :   :   :   :   :   :   :   :
707 GTTTTGTTCACCGACCGCTCAATGAATGAATGCTGCGTGCACACTCAA 658
:   :   :   :   :   :   :   :   :   :   :   :   :   :
```





```
285 rAspLeuHisIleSerSerThrProAlaAlaThrThrLeuPro..... 299
736 TTACATCTTGATAAATGATGTCCTCAACAGGTTTGTTCACCGACCGGC 687
300 .....ValMetIleLysThrGluProThrSerProThrPro 311
686 TCAATGAATGAATGTCGTGCCACTCAA...CCGGCGGATGCGGGCC 640
312 Ser.....AlaPheLysGlyProSerHisSerGlyAs 322
639 GCCGAAT.....TCATGTGTTTCGATGTTGGCAGCATTTTCAG 602
322 nProSerHisGlyThrLeuGlyLeuSerGlyThrLeuGlyAlaTyrT 339
601 ACGGCAGCTCTGGCCAGCTGCTTACACACATGGATTTCAGCTCGGTC 552
339 hrSerThrSerValProIleSerLeuSerAlaCys..... 350
551 AAACGGCTCAATACACAGCANACNTCTCTGAATCNTCNCGGCTTCITT 502
351 .....LeuAsnProAlaLeuSerGlyLeuSerSerSerThrProLe 365
501 GATCAACACACAGGCTGTGCCGNCAGCGGATTTGGTCCATCGCATGA 452
365 u..... 365
451 CGAAGATGCGACGCTCGGCATCGACGCGAGGATTTTGTGAACGGA 402
366 .....AsnGlySerAsnProLeuSerSerIleSerLeuProPro 378
401 CGGNTACGCGACGCTCCAAACCGGATTTGGATCAGATTTCNCCNNAN 352
379 HisGlySerSerThrProIleAlaProValPheThrAlaLeuProSerPh 395
351 TTCTNCCGCTTAAGTTTCCCAACGCTTCGGCGGCTAGCGTTCGAAT 302
395 eThrSerLeuThrAsnAsnPheProLeuThrGlyAsnProSerLeuAsn 412
301 CGATTTCTGTTGCTTCAACGGCAATCAGCAGCGACTGAAGTAGCGGC 252
412 roSerValSerLeuProGlySerLeuIleAlaThrSerSerThrAlaAla 428
251 TTTTCG.....CCGCGATGATGCGCGGCGATTTTG..... 222
429 ThrSerThrSerLeuProHisProSerSerThrAlaAlaValLeuSerG 445
221 .....CCTGAACNGCGCG..... 207
445 yLeuSerAlaSerAlaProValSerAlaAlaPropheProLeuAsnLeuS 462
206 .....GTAAACACCGCCCGGATNCITTTTCTCT 177
462 erThrAlaValProSerLeuPheSerValThrGlnGlyProLeuSerSer 478
176 TCAACAGCAGCTGGCTTTTTCAGCGGCATCG.....CCTTCCTTGAC 133
479 SerAsnLeuSerTyProGlyPheSerValSerAsnThrProSerValTh 495
132 TTTTCATCAGGGCGCATACCGGATATTCTTCGCCAAGCAACGCGACTT 83
495 iProAlaLeuProSerPheProGlyLeuGlnAlaProSerThrValAlaA 512
82 CGGTA.....ATGACGGGCGCGTCA 63
512 laValThrProLeuProValAlaAlaThrAlaProSer 524
seq_name: sp_fungi:094317
seq_documentation_block:
ID 094317 PRELIMINARY; PRT; 534 AA.
AC 094317;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
```

```
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 53.4 KDA SERINE-RICH PROTEIN C215.13 IN CHROMOSOME II
DE PRECURSOR.
GN SPBC215.13.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Lyne M., Rajandream M.A., Barrell B.G., Rieger M.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC -!- SIMILARITY: TO YEAST STAL.
DR EMBL; AL033534; CAA22127.1; -.
KW Hypothetical protein; Glycoprotein; Signal.
FT SIGNAL 1 14
FT CHAIN 15 534
FT CARBOHYD 31 31 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 426 426 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SEQUENCE 534 AA; 53361 MW; D62DC15C80A698E4 CRC64;
SQ
alignment_scores:
Quality: 121.00 Length: 420
Ratio: 0.582 Gaps: 15
Percent Similarity: 49.524 Percent Identity: 19.762
alignment_block:
US-09-303-518D-127/rev x 094317 ..
Align seg 1/1 to: 094317 from: 1 to: 534
1280 GGGCAGACGAAAGTCGCAAAAGCGAGGTCTCTTCGTCCTCAATTCGAAGCA 1231
67 GlyGlyThrProThrTyrTyrGlyTyrAlaThrProThrSerSerG1 83
1230 ACCCAATCCTTTCGGCGTGTGCGGTATCGCGACGATTAATTCGCGCAAA 1181
83 uProSerIlePheSerGluSerAlaThrPro.....S 94
1180 GCAGGTAGGCGAGGATGCTAGCGGCATTACGCGCTCGTAAGTACCAATC 1131
94 erGluThrAsnSerTyrSerSerProValSer..... 104
1130 GGCACCATGGCGCGGTGCGCCACCGTTGACGGCTGTCGTGAACATTGAAG 1081
105 .....SerTyrSerAspProAlaThrSerGlnLeuProSerSerTh 118
1080 TTTGTTTTTTCAGAAATGCGCGAGGTGCTAGCGGTGATGGAGTATTGT 1031
118 rSerPhePheSerPro.....ThrSerSerGluTyrThrP 130
1030 CCGGCTCGCGCGCAACCCAGCGCAACAGCTCTTCTGCTCGCGCTTCTTCG 981
130 ro.....SerSerThrGluSerSerSerLeuLeuAspProSerSer 143
980 ATAACGGAAATCTGATTGTTGTTAGCTGCCAAATAATCTGCGCGCCTTG 931
144 ValSerSer..... 146
930 TGTATTCGCGCGCTTCATACCGAACCGGAATCACGCGGTGTCTCGCT 881
147 AlaIleLeuProSerSerThrSerValGluValSerIleSerSerS 163
880 CAACCAATTCGCCGCGAGTAATTTGCGATACCTTTCGCACCCCAAAACGGTA 831
163 erLeuSerSerSer..... 167
830 CGCAAGAGCGGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 781
```





```

147 .ThrLysSerThrAlaThrSerValThrProIleProSerSerThrLeuG 163
643 GCGCG.....CCGAATTCATGCTTTTCGATGTGGCAGCATTT 606
163 lyThrThrGlyThrLeuProGluGlnThrThrProValAlaThrMet 179
605 TCA.....GACGGCAGCTGCGCGAGCTGCTTACACACATGGATTT 562
180 SerThrIleHisProSerSerThrProGluThrThrHisThrSerThrVa 196
561 ACCTCGGTCAACCGCTCAATACCCAGCANACNTCGTCTGAAATCNCN 512
196 lLeuThrThrLysAlaThrThr.....ArgAlaThrSerSer. 209
511 CGGCTCTTTGATCACAACACAGGCTGCGCGCNCAGCGATGTGTC 462
210 .....ThrSerThrProSerSerThrProGlyThrThrTrp 221
461 ATCGCATTTGACGAAGATGGCGACGCTCGGCATCGACGCGGATTTT 412
222 Ile...LeuThrGluLeuThrThrAlaAlaThrThrAlaGly..... 235
411 GCTGAACGGCAGGNTACGCGCAGCTCCACAAACG.....GATTGGA 368
236 ....ThrGlyProThrAlaThrProSerSerThrProGlyThrThrTrpI 251
367 TCAGATTGNCNANTTCGCGCGCTTAAGTTGCCAACGCTTCGGGC 318
251 leLeuThrGluLeuThrThrAlaThrThrAlaSerThrGlySer 267
317 GCGTAGGTTGCAACTCGATTGCTGCTGCTTCAACGGCA....ATCAC 271
268 .....ThrAlaThrLeuSerSerThrProGlyThrThrTrpIleLe 281
270 GACCGACTGAAGTACGCGCTTTTCG.....CGCGATGATGGCG 230
281 uThrGluProSerThrThrAlaThrValThrValProThrGlySerThrA 298
229 CGATTTTGCTGAACNGCGGGTAAACACACCGCCGCGATNCTTTTG 180
298 laThrAlaSerSerThrGlnAlaThrAlaGlyThrPro.....HisVal 312
179 TCTTCAACAGCACATGGCTTTTGTGAG.....GCATCGCCTTC 139
313 SerThrThrAlaThrThrProThrValThrSerSerLysAlaThrProSe 329
138 CTTGACTTTCATCAGGGCGCATACCGGCATATTCTTCGCCAAGCAACG 89
329 r.....SerSerProGlyThrA 335
88 CGACT.....TCGGTAATGACGGGCCCTCATATAATG 57
335 laThrAlaLeuProAlaLeuArgSerThrAlaThrThrProThrAlaThr 351
56 ACTTGCTCGGCTCGCCGCG 36
352 SerPheThrAlaIleProSer 358
seq_name: sp_archaeap_Q9HNR5

```

```

seq_documentation_block:
ID Q9HNR5 PRELIMINARY; PRT; 494 AA.
AC Q9HNR5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE VNG1983H.
GN VNG1983H.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Bailiga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leitthausser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RA "Genome sequence of Halobacterium species NRC-1."
RA Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005093; AAC20155.1; -.
KW Complete proteome.
SQ SEQUENCE 494 AA; 54739 MW; 55593A8FD31A72F3 CRC64;

```

```

alignment_scores:
Quality: 117.00 Length: 529
Ratio: 0.574 Gaps: 32
Percent Similarity: 38.563 Percent Identity: 25.520
alignment_block:
US-09-303-518D-127 x Q9HNR5 ..
Align seg 1/1 to: Q9HNR5 from: 1 to: 494

```

```

66 CGGCGCCGTCATTACCGAAGTCGCTGCTGGCGAAGATATGCCGTA 115
|||||:|||||:|||||:|||||:
20 ArgAlaArgArgArgSerArgAla..... 28
116 TGCGCCCTNGATCAAGTCAAGGAGCGATGCCGTCAAAAAAGGCCAA 165
:|||||:|||||:|||||:
29 .....GluArgAspProGlyArgArg.....G 36
166 GTGCTGTTTGAAGACAAAAGATCCGGCGTGTGTACCGCGCGNGT 215
:|||||:|||||:|||||:
36 lyAlaValLeuGlyAlaArgThrHisAspArgGlyGlyValArgAla... 51
216 TTCAGGCAAAATCCCGCCCAT.....CCATCGGCGGCAAAAGCGGTAC 259
|||||:|||||:|||||:
52 ...ArgArgGlyArgArgHisArgProProProLeuArgThrAlaArgG 67
|||||:|||||:|||||:
260 TTCAGTCGTCGTGATTGCCGTTCAAGGCCAA.....CGAC 294
:|||||:|||||:|||||:
67 yArgThrGlyValAspSerAlaAlaArgAspAlaAlaProValProA 84
|||||:|||||:
295 GAAATCGAGTTCCA..... 308
|||||:
84 rgThrAlaValArgAspGlyGlyGlnThrArgGluCysProSerArg 100
309 .....ACGCTAGCGCGCGCAAG 325
|||||:|||||:
101 GlyProValSerValGlyValThrAlaGluProProValArgArgPr 117
326 CGTTGGCAAACTTAAGCGCGGANGA..... 350
|||||:|||||:
117 oValGlyArgLeuLeuArgValLeuArgGlyArgAlaArgValAspA 134
350 ..... 350
134 rgAlaGlyAlaValArgGlyArgAlaArgAlaAspThrAlaSerGlyArg 150
351 .....ANTNNGNCAATCTGATCCAAATCCGGTTT...GTGACTGCCG 391
151 ProProArgArgGlyGluThrArgProHisArgAlaAspValGlyCysG 167
392 T.....GCGTANCCGTCGG 405
167 yArgProValCysAspArgValGlyAlaAsnArgProAlaGluProArgH 184
406 TTCAGCAAAATCCCTCCGTCGATGCCGAGCGGTTCGCCATCTT..... 449
:|||||:|||||:

```





940 CACGATTATTTGGGACGCTACCAACATCAGATTTCGTTTATCAGAAAGG 989  
 480 rSerArgSerArgThrProProValThr.....ArgArgArg 492  
 990 CCGCAGCAAAAGACTGTCGGCTGGG.TTGCCTGGCAGCGGACGACAAATAC 1038  
 493 SerArgSerArgThrProProValThrArgArgSerArgSerArgSerArg 509  
 1039 TCCA.....TCACGCGTACGACCTCGGCATTTCT 1070  
 509 rSerProValThrArgArgSerArgSerArgThrSerProValThrA 526  
 1071 GAAACAACTCTCTCAAGTTCACGACGCGTCAACGGTGGCGACCGG 1120  
 526 rArgArgSerArgSerArgThrSerProValThr.ArgArgArgSerAr 542  
 1121 CCATGGTGGCGATGGTGTACTTACGACGCGTAAATGCGCTAGACATCCTG 1170  
 542 gSerArgThrProProAlaileArgArgSerArgSerArgSerArgThrProL 559  
 1171 CCTACCTGCTTTTTCGCGATTAAATCGTCGCGATACCGACACGCGCA 1220  
 559 euLeuPro.....ArgLysArgSerArgSerArgSer 569  
 1221 AGCATTTGGTGGTGGTGGATGGACGAAGAAGACCTCGC 1259  
 570 ProLeuAlaile.....ArgArgArgSerArg 578  
 seq\_name: sp\_human:O15038  
 seq\_documentation\_block:  
 ID O15038 PRELIMINARY; PRT; 1783 AA.  
 AC O15038;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE KIAA0324 PROTEIN (FRAGMENT).  
 GN KIAA0324.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=97349984; PubMed=9205841;  
 RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyatake A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro."  
 RL DNA Res. 4:141-150(1997).  
 DR EMBL; AB02322; BAA20782.2; -.  
 FT NON\_TER 1  
 SQ SEQUENCE 1783 AA; 190940 MW; 660302F6FD4179AB CRC64;  
 alignment\_scores:  
 Quality: 116.50 Length: 447  
 Ratio: 0.613 Gaps: 19  
 Percent Similarity: 42.506 Percent Identity: 24.161  
 alignment\_block:  
 US-09-303-518D-127 x O15038 ..  
 Align seg 1/1 to: O15038 from: 1 to: 1783  
 12 CAAAAAAGGTCTAAACCTGCCATCGCGGACGACCGGACGACGATTT 61  
 748 ArgSerArgThrProProAlaileArgArgSerProSerValSerSerPr 764  
 62 ATGACGGCGCGTCATTACCGAAGTTCGGTTCGTTGGCGAAGATATGCC 111



```
764 oGluProAlaGlu.....LysSerArgSerSerArgArgArgSera 779
112 GGTATGCGCCCTNGATCAAGTCAAGGAAGCGCATCGCTCAAAAAGG 161
779 laSerSerProArgThrLysThrSerArgArgGlyArgSerProSer 795
162 CCAAGTCTGTTTGAAGACAAAAAGNATCGGCGCTGTTGTTACCGCGC 211
796 ProLysPro.....ArgGlyLeuGlnArgSe 804
212 CNGTTCAGCAAAATCGCGCCCATCCATCGCGCGAAAGCGCTACTT 261
804 r..... 804
262 CAGTCGGTCGTGATTGCCGTGAAGCAACGAGCAAAATCGAGTTGCAACG 311
805 .....ArgSerArgSerArgGlnLysThrArgThr 815
312 CTACGCGCCGAAGCGTTGGCAAACTTAAGCGCGCGANGAANTNNNGCA 361
816 ThrArgArgArgAsp.....ArgSerGlySerSerGln 826
362 ATCTGATCAAAATCGCGTTTGTGGACTGCGCTCCGTANCCGTCGCTCAGC 411
826 nSerThrSerArgArg.....ArgG 833
412 ARAATCCCTGCGCTGATGCGCGAGCGCTTCGCCATCTTCGTAATGCGAT 461
833 InArgSerArgSerArgValThrArgArgArgGlyGlySer 849
462 GGACACCAATCGCTNGC..... 479
850 GlyTyrHisSerArgSerProAlaArgGlnLysSerSerArgThrSerSe 866
480 .....GCCAGA.....CCCTGTGGTGTGATCAAAAGA 507
866 rArgArgArgArgGlyArgSerArgThrProProThrSerArgLysArgS 883
508 GCGGCGGANGATTTCAGACGANGNTGCTGTATTGACCGCTTGACCGGA 557
883 erArgSer.....ArgThrSerProAla.....ProTyrLysArg 894
558 CGGTAAATCATGTGTGAAGCAGCTGGCGCAGAGCTGCCGTCTGAAA 607
895 Ser.....ArgSerArgAla.....Se 900
608 ATGTGCGCAACATGAAACATCAATTCGCGCGCGCGATCGCGCGCT 657
900 rProAlaThrHisArgArgSerArgThrProLeuLysSerArg. 916
658 TTGAGTGGCAGCACATTCATTTCATTGAGCGCGCTGCGTGCAACAAAC 707
917 .....ArgArgSerArgSer 921
708 CGTTTGGACCATCAATATCAAGATTAATTCGATCGGACGTTGTTG 757
922 ArgThrSerProValSerArgArgArgSerArgSerArgSerValTh 938
758 CACAGCGCTGTGAACACGAGCGCGCTGATTGCTTTGGGTGTTCTCAA 807
938 rArgArgArgSerArgSerArgAlaSerProValSerArgArgArgSera 955
808 GTCAACAACACCGCTCTTGGTACCGTTTGGTGGCAAGTATCGCA 857
955 rSerArgThrProProValThr..... 962
858 AATTACTGCGGCGAATGTTGGTACGACGACAGCAACCGCTGATTTCGGTT 907
963 .....ArgArgArgSerArgSerArgThrPr 971
908 CGGTATTGAACGCGCGAT.....TACACAAGCGCG 939
971 oThrThrArgArgArgSerArgThrProProValThrArgArga 988
```

```
940 CACGATTATTGGGACGCTACCAATACAGATTTCGTTATCGAAGAGG 989
988 rgSerArgSerArgThrProProValThr.....ArgArgArg 1000
990 CCGCAGCAAAAGAGCTGTTGGCTGGG.TTGGCGCGCAGCCGGAACAATAC 1038
1001 SerArgSerArgThrSerProLysThrArgArgSerArgSerArgTh 1017
1039 TCCA.....TCACGCGCTACGACCTCGCGCATTTCT 1070
1017 rSerProValThrArgArgArgSerArgThrSerProValThr 1034
1071 GAAAAACAACACTTCAAGTTCAACGACGCGCTCAACGCTGGCGACCGC 1120
1034 rgArgSerArgSerArgThrSerProValThr.ArgArgArgSerAr 1050
1121 CCATGTCGCGGATTTGCTACTTACGAGCGGTAATGCCGCTAGACATCTG 1170
1050 gSerArgThrProProAlaLysArgArgSerArgSerArgThrProL 1067
1171 CCTACCGCTCTTTTCGCGGATTTAATCGTCGCGGATACCGACGCGCA 1220
1067 euLeuPro.....ArgLysArgSerArgSerArgSer 1077
1221 ACATTTGGTGTGCTTGAATTTGGAGAGAAAGACCTCGC 1259
1078 ProLeuAlaLys.....ArgArgArgSerArg 1086
```

seq\_name: sp\_human:060382

seq\_documentation\_block:

```
ID 060382 PRELIMINARY; PRT; 1791 AA.
AC 060382;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE K1AA0324 (FRAGMENT).
GN K1AA0324.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,
RA Goodwin L., Bryant J., Tesmer J., Meincke L., Longmire J., White S.,
RA Teng S., Tatum O., Campbell C., Fawcett J., Deaven L.;
RT "Sequencing of Human Chromosome 16p13.3";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ricke D.O.;
RT "Large Scale Sequence Analysis and Annotation with the Sequence
RT Comparison Analysis (SCAN) System.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004493; AAC08453.1;
FT NON_TER 1
SQ SEQUENCE 1791 AA; 191306 MW; 3A7B5530AE95F3E CRC64;
```

alignment\_scores:

Quality:	116.50	Length:	447
Ratio:	0.613	Gaps:	19
Percent Similarity:	42.506	Percent Identity:	24.161

alignment\_block:

US-09-303-518d-127 x 060382 ..

Align seg 1/1 to: 060382 from: 1 to: 1791

12 CAAAAAAGCTCTAAACCTGCCATCGCGGCGACCGGCAAGTCATTT 61

964 .....ArgArgArgSerArgSerArgThrPr 972  
908 CGGTATTGAACGGCCGAT.....TACACAGCGCG 939  
972 othrThrArgArgSerArgSerArgThrProProValThrArgArg 989  
940 CAGGATTATTGGGACGCTACCAATCAGATTTCGGTTATCGAAGAGG 989  
989 rgSerArgSerArgThrProProValThr.....ArgArg 1001  
990 CGCAGCAAGAGCTGTCGGCTGGG.TTGGCCCGCAGCGGACCAATAC 1038  
1002 SerArgSerArgThrProProValThrArgArgSerArgSerArgTh 1018  
1039 TCCA.....TCAAGCGGTACGACCTCGGCCATTTCCT 1070  
1018 rSerProValThrArgArgSerArgSerArgThrSerProValThr 1035  
1071 GAAAAACAACCTTCAAGTTCAAGACAGCCCTCAACGGTGGCGACCGG 1120  
1035 rgArgArgSerArgSerArgThrSerProValThr.ArgArgArgSerAr 1051  
1121 CCATGGTGGGATGCTTACTTACGAGCGGTAAATCGCGCTAGACATCCTG 1170  
1051 gSerArgThrProProAlaIleArgArgSerArgSerArgThrProL 1068  
1171 CCTACCTGCTTTTGGCGGATTTAATCGTGGCGGATACGACAGCGCGA 1220  
1068 euLeuPro.....ArgLysArgSerArgSerArgSer 1078  
1221 AGCATGGGTGCTTGAATTGGACGAGAAAGACCTCGC 1259  
1079 ProLeuAlaIle.....ArgArgArgSerArg 1087

seq\_name: sp\_human:Q9UQ35

seq\_documentation\_block:  
ID Q9UQ35 PRELIMINARY; PRT: 2752 AA.  
AC Q9UQ35;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE RNA BINDING PROTEIN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ohtaki S., Umeki K., Sawada Y.;  
RT "Homo sapiens mRNA for RNA binding protein, complete cds.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB016092; BAA83718.1; -;  
DR InterPro: IPR002965; P-rich\_extensn.  
DR PRINTS: PR01217; PRICHEXTENS.  
SQ SEQUENCE 2752 AA; 299672 MW; 109C64F181097123 CRC64;

alignment\_scores:  
Quality: 116.50 Length: 447  
Ratio: 0.613 Gaps: 19  
Percent Similarity: 42.506 Percent Identity: 24.161  
alignment\_block:  
US-09-303-518D-127 x Q9UQ35 ..  
Align seg 1/1 to: Q9UQ35 from: 1 to: 2752  
12 CAAAAAGTCTAAACCTCCCATCGCGGACAGCCGAGCAAGTCAATT 61  
1717 ArgSerArgThrProProAlaIleArgArgSerProSerValSerPr 1733  
62 ATGACGGGCGCGTCATTTACCGAAGTCGCGTTCGTTGGCGAAGATATGCC 111

749 ArgSerArgThrProProAlaIleArgArgSerProSerValSerPr 765  
62 ATGACGGGCGCGTCATTTACCGAAGTCGCGTTCGTTGGCGAAGTATGCC 111  
765 oGluProAlaGlu.....LysSerArgSerArgArgArgSerA 780  
112 GGTATGCGCCCTGATGAAGTCAAGGAAGCGGATGCGCTCAAAAAAGG 161  
780 laSerSerProArgThrLysThrThrSerArgArgGlyArgSerProSer 796  
162 CCAAGTCTCTTTGAAGACAAAAAGNATCCGGCGGTGGTTCACCGCGC 211  
797 ProLysPro.....ArgGlyLeuGlnArgSe 805  
212 CNGTTTCAGCAAAATCGCGCCATCCATCGCGGCGGAAAAAGCGCTACTT 261  
805 r..... 805  
262 CAGTCGGTCGTGATTCGCTTGAAGCAACGAGCAATCGATTGGAAGC 311  
806 .....ArgSerArgSerArgGluLysThrArgThr 816  
312 CTAGCGCCGCAAGCGTTGCAACITTAAGCGCGGANGAANTNNGNCA 361  
817 ThrArgArgArg.....ArgSerGlySerSergl 827  
362 ATCTGATCCAATCGCGTTTGTGGACTCGCTGCGTANCCGTCGTTTCAGC 411  
827 nSerThrSerArg.....Arg 834  
412 AAATCCCTCGCTGATGCGGACCGCGCTTCGCATCTTCGTCAATCGAT 461  
834 InArgSerArgSerArgValThrArgArgArgGlyGlySer 850  
462 GGACCAATCCGCTNGC..... 479  
851 GlyThrHisSerArgSerProAlaGlnGluSerSerArgThrSerSe 867  
480 .....GGCAGA.....CCTGTGGTTGTGATCAAGAA 507  
867 rArgArgArgGlyArgSerArgThrProProThrSerArgLysArgS 884  
508 GCGGCGGANGATTCAGACGANGNTGCTGTTATTTAGCCGCTTTGACCGA 557  
884 eArgSer.....ArgThrSerProAla.....ProTrpLysArg 895  
558 GCGTAATCCATGCTGTAGGCGAGCTGGCGCAGAGTGGCGTCTGAAA 607  
896 Ser.....ArgSerArgAla.....Se 901  
608 ATGCTGCCAATCGAAACACATGAATTCGGCGCGCCCGCATCCGCGCGGT 657  
901 rProAlaThrHisArgArgSerArgSerArgThrProLeuLysSerArg. 917  
658 TTGAGTGGCAGCGCACATTCATTTCAATTGAGCGGTGCTGCAACAAAC 707  
918 .....ArgArgSerArgSer 922  
708 CGTTTGGACCATCAATATCAAGATGATATGCCATCGGAGCTTTGTTTG 757  
923 ArgThrSerProValSerArgArgSerArgSerArgThrSerValTh 939  
758 CAACAGCGCTCTGAACACCGGCGGTGATTCGTTGGTGGTCTCAAA 807  
939 rArgArgArgSerArgSerArgAlaSerProValSerArgArgArgSerA 956  
808 GTCAACAAACACCGCTCTTCGCTACCGCTTTTGGTGGCGAAGATATGCCA 857  
956 rgSerArgThrProProValThr..... 963  
858 AATTACTGCGGGCAATTGGTTCACGACAGCAACCGCGTATTCGCGTT 907

```

1733 oGluProAlaGlu.....LysSerArgSerArgArgArgSera 1748
112 GGTATGCGCCCTNGATGAAGTCAAGAAAGGCGATCCGTCACAAAGG 161
1748 laSerSerProArgThrLysThrSerArgArgGlyArgSerProSer 1764
162 CCAAGTGTCTGTTTGAACACAAAAGNATCCGGCGTGGTGTACCGCC 211
1765 ProLysPro.....ArgGlyLeuGlnArgSe 1773
212 CNGTTTCAGCAAAATCGCGCCATCCATCGCGGCAAAAGCGCTACTT 261
1773 f..... 1773
262 CAGTGGTGTGATTGCGTGAAGCAAGCAAGCAAGTTCGAGTTCGAACG 311
1774 .....ArgSerArgSerArgGlyArgGlyThrArgThr 1784
312 CTACGCGCCGCAAGCCTTGCAAACTTAAGCGCGCGANGAANTNNGCA 361
1785 ThrArgArgAsp.....ArgSerGlySerSerG1 1795
362 ATCTGATCAATCCGCTTTGTGACTGCGCTCGGTANCCGTCCTCAGC 411
1795 nSerThrSerArgArg.....ArgG 1802
412 AAAATCCCTGCGTGCATGCGAGCGCGTTCGCCATCTCGCAATCGAT 461
1802 InArgSerArgSerArgSerArgValThrArgArgArgGlyGlySer 1818
462 GGACACCAATCCGCTNGC..... 479
1819 GlyTyrHisSerArgSerProAlaArgGlnGluSerSerArgThrSerSe 1835
480 .....GCACAG.....CCTGTGGTGTGATCAAGAA 507
1835 rArgArgArgArgGlyArgSerArgThrProThrSerArgLysArgS 1852
508 GCCGCGCAGGATTTTCACAGCANGTCTCGTATGAGCGGTTTGACCA 557
1852 erArgSer.....ArgThrSerProAla.....ProTrpLysArg 1863
558 CCGTAATCAATGATGTGTAAAGCAGCTGCGCAGACGTCCGCTCGAAA 607
1864 Ser.....ArgSerArgAla.....Se 1869
608 ATGCTGCCAATCAACACATGAATTCGGCGGCCCGCATCCGCGCGT 657
1869 rProAlaThrHisArgArgSerArgSerArgThrProLeuIleSerArg 1885
658 TTGAGTGGCAGCACATTCATTTCATGAGCGCGTCCGTCGCAAAAC 707
1886 .....ArgArgSerArgSer 1890
708 CGTTTGGACCAATTCATCAGATGTAATGCCATCGACGTTTGTGTTG 757
1891 ArgThrSerProValSerArgArgArgSerArgSerArgThrSerValTh 1907
758 CAACAGCGCGTCTGAACACCGAGCGCGTATCTTGGTGGTGTCTCA 807
1907 rArgArgArgSerArgSerArgAlaSerProValSerArgArgSerA 1924
808 GTCAACAACACCGCTCTGCGTACCGTTTGGTGGCGAAAGTATCGCA 857
1924 rgSerArgThrProProValThr..... 1931
858 AATTACTCGCGCGAATTGGTTGACGACAGCAACCGGTGATTTTCGGTT 907
1932 .....ArgArgSerArgSerArgThrPr 1940
908 CGGTATTGAAGCGCGAT.....TACACAAAGCGCG 939
:|||||:

```

```

1940 oThrArgArgArgSerArgSerArgThrProProValThrArgArgA 1957
940 CACGATTATTGGACGCGCTACCAATCAGATTCCTGTTATCGAAGAG 989
1957 rgSerArgSerArgThrProProValThr.....ArgArgArg 1969
990 CCGCAGCAAGAGCTGTTCGGCTGGG.TTGGCGCGCAGCGGACAAATAC 1038
1970 SerArgSerArgThrSerProIleThrArgArgSerArgSerArgTh 1986
1039 TCCA.....TCACGCGTACGACCTCGGCCATTCTCT 1070
1986 rSerProValThrArgArgSerArgSerArgThrSerProValThra 2003
1071 GAAAAACAACCTCTCAAGTTACGAGCGCTCAACGCTGGCGACCGCG 1120
2003 rgArgArgSerArgSerArgThrSerProValThr.ArgArgArgSerAr 2019
1121 CCATGTGCGCGATTGCTACTTACGAGCGGTAAATCCCGCTAGACATCTG 1170
2019 gSerArgThrProAlaIleArgArgSerArgSerArgThrProL 2036
1171 CTCTACCTGCTTTTTCGGCGATTAAATCGTCGGGATACCGACACGCGCA 1220
2036 euLeuPro.....ArgLysArgSerArgSerArgSer 2046
1221 AGCATTTGGTGTCTTGAATTTGACGAGAGACGCTCGC 1259
2047 ProLeuAlaIle.....ArgLysArgSerArg 2055
seq_name: sp_invertebrate:0900V2
seq_documentation_block:
ID Q9U0V2 PRELIMINARY; PRT: 2207 AA.
AC Q9U0V2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE POSSIBLE MUS308 HOMOLOG (FRAGMENT).
GN L8342.01.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Masuy D., Purnelle B., Goffeau A., Ivens A.C., Lawson D., Quail M.,
RA Rajandream M.A., Barrell B.G.
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL: AL122012; CAS58415.1;
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00270; DEAD; 1.
DR Pfam: SM00490; HELIC_C; 1.
DR SMART; SM00490; HELIC_C; 1.
KW ATP-binding; Helicase.
FT NON_TER 2207 2207
SQ SEQUENCE 2207 AA; 229262 MW; E461565DA0017B0C CRC64;

```

```

alignment_scores:
Quality: 116.00 Length: 447
Ratio: 0.611 Gaps: 21
Percent Similarity: 42.506 Percent Identity: 22.819

```

Align seg 1/1 to: Q9U0V2 from: 1 to: 2207

1313 TTACGCAACAGCGGCCNTATTCGTTTGGCTTAT

2000

1313 TTACGCAACAGCGGGCCNTATTCGTATTGCCCGGCAGACGAAGCTGCA 1204

```

569 ProlAlpheGlyPheGlyIleAsnSerValSerSerSerVal...Se 584
378 ACCGGATTGGATCAGATTGCNNCNANNTCTCCGCCGTAAAGTTTGGC. 330
; : ::::::::::::::::::::|::|::|::|::|::|::|::|::|::|::|
584 rThrThrThrSerThrAlaTrhAlaAIsrSrlnProPhLeuPheGlyA 601
329 ....AACGTCCTGGCGCGTAGCTTGCAACTCGATTTCGTCGCTG.. 288
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
601 laProGlnAlaserAlaAIsrPheThrProAlaMetGlySerIlePhe 617
287 .....CCTCAACGGCAATCAGACCAGCTGAAGTACGCGCTT 250
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
618 GlnPheGlyLysProProAlaLeuProThrThrThrValThrPh 634
249 TTGCCCCGATGGATGCGCGCATTTTGCTGTAACNGCGCGGTAAACA 200
||||::|::|::|::|::|::|::|::|::|::|::|::|::|
634 eSerGlnSerLeuHisThrAlaValProThrAlaThrSerSerAlaA 651
199 CCAGCGCCGATNCTTTTTGCTTCAACAGCACACT...TGGCCCTTTT 153
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
651 IaaspPheSerGlyPheGlySerThrLeuAlaThrSerAlaProAla 667
152 ACGGCATCGCTTCCTTGACTTTTCATCNACGGGGCGCATACGGCAT 103
; : ::|::|::|::|::|::|::|::|::|::|::|::|::|::|
668 SerSerGlnProThrLeuThrPheSerAsnThrSerThrProThrP 684
102 TTCGCCAAGCAACGCGACTTCGGTAATGACGGCGCGTCATAAATG 53
|||::|::|::|::|::|::|::|::|::|::|::|::|::|
684 nilePropheGlySerSerAlalylsserProLeuProSertyrPr 701
52 GCTCCGGTCTGCCCGGATGGGC 30
::|::|::|::|::|::|::|::|::|::|::|::|::|
701 IaasnProgInProAlaPheGly 708

seq_name: sp_todent:Q99K60

seq_annotation_block:
ID Q99K60 PRELIMINARY; PRT; 631 AA.
AC Q99K60;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MUCIN 1, TRANSMEMBRANE.
GN MUC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleo
OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae
NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC005441; AH05441.1; -.
DR MGD; MGI:97231; Muc1.
DR InterPro; IPR000082; SEA.
DR Pfam; PF01390; SEA; 1
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS50024; SEA; 1.
KW Transmembrane.
QQ SEQUENCE 631 AA; 64690 MW; B4236DF66655F4C3 CRC64;
```

```

alignment_scores:
  Quality: 112.00
  Ratio: 0.563
  Percent Similarity: 42.521
  Length: 468
  Gaps: 17
  Percent Identity: 22.436

alignment_block:
  US-09-303-518D-127/rev x Q99K60

```

Align seg 1/1 to: Q99K60 from: 1 to: 631

1343 CAGCCTTCCTTCTCAANGGTTTCCAGCACCTTA.....CG 1309

```

111 ... |||.....:|||||
35 GlAspThrSerSerSerLeuAlaSerThrThrProValHisSerSe 51
1308 CAACAGCGGCGCCTATTCGTAATTCGCGCGCAGCAAGCTGCACAAAG 1259
||||| ||| ... |||.....:|||||
51 rAsnSerAspProAlaThrArgProGlyAspSer..... 63
1258 CGAGGTCTCTTCGTCCTCAATTCACCAAGCAACCAATGCTTCGCGCTGTCG 1209
...
64 .....:Thr 64
1208 GTATCGCCGACGATTAAATCGCGCAAAAGAGGTAGCAGGATGCTAG 1159
|||||: |||.....:|||||
65 SerSerProValGlnSerSerThrSerProAlaThrArgAlaProG1 81
1158 CGCATTCAGCGCTCGTGAAGTCAATCGGACCATCGCGGGTGGCCAC 1109
...
81 uAspSerThrSerThrAlaValLeuSerGlyThr.....SerSerProA 96
1108 CGTTGACGGCTGCTGTGAATTCGAGAGTTGTTTTCAGGAATGGCGG 1059
||||| ||| |||
96 laThrThrAlaProValAsnSerAlaSer..... 105
1058 AGGTCGTACGCGTGTAGTATTCGCGGCTCGCGGCAACCCAGCC 1009
...
106 .....:SerProValAlaHisGlyAspThrSerSerPr 116
1008 GAACAGCTCTTTCGCGGCTCTTCGATAACGGAAATCTGATGTGGT 959
| .....: |||
116 oAlaThrSerLeuSerLysAspSer..... 124
958 AGGTCGCAAAATATCGTGGCGCTGTGTAATCGCGCGTCAATACC 909
...
125 .....:AsnSerSerProValAlaHisSerGlyThrSerSer 136
908 GAACCGGAAATCATCGGCTGTGCGTCAACCAATTCGCGCGCATAT 859
||| ||| .....:|||||
137 AlaproAlaThrAlaProValAspSerThrSerProValValHi 153
858 TTGGATCTATTCGACCCCAACAGGTACGAGAGCGTGTGTTGTA 809
|||||: ||| |||
153 sglyGlyThrLeuSerProAlaThr.....SerProGlyAspSer 168
808 CTTGAGAACCAACCAAGCAATCAGCGCTCGGTGTTTCAGAGCGCTGT 759
||| .....: ||| |||
168 hrSerSerProAspHisSerThrSerSer.....Pro 179
758 GCAACAAACGTCGATGGCAATTCATCTGATAATGATGTCACAAAC 709
|||||: ||| |||
180 AlathrArgAlaProGluAspSerThrSer.....Th 190
708 GGTGTTGTTCCAGCGGCTCAATGAATGAATGTCGCTGCTCA 659
| .....: ||| |||
190 rAlaValLeuSerGlyThrSerSer..... 198
658 RACCGCGGATCGCGCGCGGCAATTCATGTTTCGATGTTGCGCA... 612
||||| ||| .....: |||
199 ...ProAlaThrAlaProValAspSerThrSerProValAlaHis 214
612 ..... 612
215 AspAspThrSerSerProAlaThrSerProSerGluAspSerAlaSerSe 231
611 ....GCATTTTCAGCGCTGTCGCGCAGTGCCTTACACACATGGA 566
...
231 rProValValHisGlyThrSerSerProAlaThr..... 243
565 TTTCAGCTCGCTCAACGGCTCAATACCAACANACNCTGCTCAATCN 516
||| ||| .....: |||
244 .....SerProLeuArgAspSerThrSerProValHisSerSer 257
515 TCNCGGCTCTTTCATCACACACAGGCTGTCGCGCAGC..... 474
: ||| ||| .....: |||

```

```

258 AlaserIleGlnAsnIleLysThrThrSerAspLeuAlaSerThrProAs 274
473 ...GGATTGGTTCATTCGATTCGAGAGATGCGGACGCGCTCGCAT 428
... |||.....: ||| |||
274 pHisAsnGlyThrSerValThrThrThrSerAlaLeuGlySerAlat 291
CGAGCGCAGGAGTTTCTGTAACGCGACGNTACGACGCGCAGTCCACAAA 378
...
291 hrSer..... 292
377 CGGATTCGATCAGATTGCNNCANNNTTCGCGCTTAAGTTTGCAA 328
|||||
293 ProAsp.....HisSerGlyThrSerThrThrThrAsnSerSe 305
327 CGCTTCGCGCGTAGCTTCGAACTCGATTTCGTCGTCGCT...TCAA 281
... ||| .....: ||| |||
305 rGluSerAspLeuAlaThrThrProValThrSerSerMetProPheSer 322
280 CGGCAATCACGACGACTGAGTAGCGGCTTTTCGCGCGCATGGATGCG 231
|||||: ||| .....: |||
322 hrThrLysValThrSerGlySerAlaIleIleProAspHisAsnGlySer 338
230 GCGATTTCGCT..... 219
339 SerValLeuProThrSerSerValLeuGlySerAlaThrSerLeuValTy 355
218 .GAACNGCGCGGTAACACACCGCGGATNCTTTTGTCTTCAACA 170
... ||| .....: ||| |||
355 rAsnThrSerAlaIleAlaThrPro.....ValSerAsnG 368
169 GCACTTGGCTTTTTCGCGGATCGCTTCTGACTTTCATCNCAGGG 120
||||| ||| .....: |||
368 lyThrGlnProSerValProSerGlnThrProValSerProThrMetAla 384
119 CGCATACCGCATATTCGCAACGACGCGACTTCGCGTAATGACGG 70
... ||| .....: |||
385 ThrThrSerSerHisThrThrAlaSerSerSerTyrThrVa 401
69 CCGG 66
|||
401 lPro 402
seq_name: sp_plant:09SUVO
seq_documentation_block:
ID 09SUVO PRELIMINARY; PRT; 857 AA.
AC 09SUVO;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 96.9 KDA PROTEIN.
GN F8B4.120 OR A14G32420.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID:3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,
De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
Villaroel R., Gielen J., Van Montagu M., Hoheisel J., Mewes H.W.,
Mayer K.F.X., Lemcke K., Schueller C.;
RA Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RL [2]
RN SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RL [3]
RN SEQUENCE FROM N.A.
RA Terry N., Ardiles W., Buysshaert C., Dasseville R., De Clerck R.,
De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villaroel R.,
Gielen J., Van Montagu M., Mewes H.W., Lemcke K., Mayer K.F.X.;
RA

```









SQ SEQUENCE 529 AA; 58716 MW; 5DC1216237DF772D CRC64;

alignment\_scores:  
Quality: 108.00 Length: 473  
Ratio: 0.548 Gaps: 24  
Percent similarity: 41.649 Percent Identity: 24.101

alignment\_block:  
US-09-303-518D-127 x P94909 ..

Align seg 1/1 to: P94909 from: 1 to: 529

27 CCTGCCCATCGGGCAGACGAGCAAGTATTTATGACGGGCGCTCA 76  
:::|||||  
128 ProAlaArgGlyValLeuArgHisArgHisArgAlaArgGl 144  
:::|||||  
77 TTACCGAAGTCGGTGCTTGGCGAGAATAATGCCGTATCGCCCTNG 126  
|||||||  
144 yLuarGlyArgGlyProArgGlnGlnValProArgGlnHisPro... 159  
:::|||||  
127 ATGAAGCTCAAGGAGCGCATGCCGTCFAAAAAGCCAAGTCGTGTTGA 176  
:::|||||  
160 .....ArgGlyArgAspArgAlaGlyArgProGlyLeuHisArg 173  
:::|||||  
177 AGACAARAAGNATCC.....GGCGCTGGTGTATTACCGCGCNG 214  
|||||||  
174 ArgArgArgArgAlaArgArgArgGlnGlyArgProGlnValArgHisGl 190  
:::|||||  
215 TTTCAGGCAAAATCGCGGCATCCATCGGGGGAAGCGGTACTTCAG 264  
|||||||  
190 yasPaspGlnHisArgAlaaspProArgProArgaspProArgAlag 207  
:::|||||  
265 TCGGTCGTGATTCGCTGTAAGGCAACGAGAAATCGAGTTCGAACGCTA 314  
:::|||||  
207 lyHisHisProLeuArgGluGlyGlyGlnGlyGlnGlyGlnAlaArgProleu 223  
:::|||||  
315 CCGCGCCCAAGC.....GTTGGCAACTTAAAGCGCGGANG 349  
|||||||  
224 ArgAlaArgGlyGlyGlnGlyGlyProGlyGlyProAlaGlyArgHisPr 240  
:::|||||  
350 AANTNNNGCAATCTGNATCCATCCGTTGTGGAGTCCGCTGGGTANC 399  
:::|||||  
240 oAlaLeuGlyaspGluasp.....ArgP 248  
:::|||||  
400 CGTCGCTTCAGCAAATCCCTGCGTCGATCCGACGCCGCTTCGCATCTT 449  
|| :::|||||  
248 roArgAlaArgGlnGlyAlaArgAlaAlaHisArgGluAlaLeuHisLeu 264  
|| :::|||||  
450 CGTCAATGGATGGACACCATCCGCTNGCGGAGACCCTGGTGTGT... 497  
|||  
265 ArgLeu.....GlnArgArgGlyArgAlaaspGlyCysar 277  
|||  
498 ....GATCAAGAAGCCGNCANGATTCAGACGANGTNTCTGTPATTG 543  
:::|||||  
277 gAlaGlnGlyArgAlaArgArgAlaArgAlaGlyGlyGlyLeu 293  
:::|||||  
544 AGCGGTTTGACCGAGCGTAAATCATCATGTGTATGAGCACGCTCGCCAGA 593  
|||  
294 ProArgGluasparg..... 299  
594 CGTGGCGCTGTAATGCTGCCAACATCGCAACACATGAATTCGGCGGCC 643  
:::|||||  
300 .....ValArgAlaHisarg.....P 305  
644 CGCATCCGGCGGTTGAGTGGCAGCACATCATTTTCATGACCGGCTC 693  
|||||||  
305 roArgProArgGlyArgGlyGlyAlaAlaArgLeuaspArgAlaGlyGl 321  
694 GGTCGAACAACACCCCTTTGGACCATCAATATCAAGATGTAAATGGCAT 743  
:::|||||  
321 yValGlyAlaArgProAlaArgProArgLeuArgHisProArgProA 338

RC STRAIN-BREED ANGUS;  
 RA MEDLINE=2023253; PubMed=10759843;  
 RT Jhang W., Gupta D., Gallagher D., Davis S., Bhavanandan V.P.;  
 "The central domain of bovine submaxillary mucin consists of over 50  
 tandem repeats of 329 amino acids: chromosomal localization of the  
 BSM1 gene and relations to ovine and porcine counterparts.";  
 RL Eur. J. Biochem. 267:2208-2217(2000).  
 DR EMBL; AF178428; AAF67279.1; -;  
 FT NON\_TER 1 1  
 FT NON\_TER 818 818  
 SQ SEQUENCE 818 AA; 73014 MW; B5BB44F84F6F86B CRC64;

alignment\_scores:  
 Quality: 108.00 Length: 431  
 Ratio: 0.535 Gaps: 25  
 Percent Similarity: 46.868 Percent Identity: 24.594

alignment\_block:

US-09-303-518d-127/rev x Q9N1P0 ..

Align seg 1/1 to: Q9N1P0 from: 1 to: 818

```

1301 GGGCCNTATTCGCTATTTGCC...GGCAGACGAAGCTGCACAAAGCGAG 1255
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
261 GlyProGlyThrSerAlaProSerGlyGluThrSer.....GluSerAr 275

1254 GTCCTCT.....TCGTCCAATTCACCAAGCAACCAATGCTTGGC 1217
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
275 gSerSerValProGlySerGluThrThrGlnGlnProGlyAlaGly. 291

1216 CGCTGTCGCTATCGCGACGATTAATTCGCGCAAAAGCAGGTAGGCAGG 1167
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
292 .....SerGluSerProThr..... 296

1166 ATGCTAGCGCATTCAGCGCTCGTAAGTACCATCGGACCAATCGGCGCG 1117
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
297 LeuSerProGlyValThrArgThrThrAlaLeuArgGlySerGluThrAr 313

1116 GTCGCCACCGTTCACGCGCTGCTGTAACCTTGAAAGATTGTTTTCAGCA 1067
| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
313 gValProSer...ThrGlyValSerGlyLeuPro..... 323

1066 AATGGCGGAGGTCGTACGCGTGATGGAGTATTTGCCGCGTCGCGCGCA 1017
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
324 ..... 325

1016 ACCAGCGCAACAGCTCTTTGCTGCGGCTCTCTGATAACGGAATCTG 967
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
326 ThrGlnGlyGlySerAlaAlaThrGlyGlySerGlyAla..... 338

966 ATGTGGTAGCTCCCAATAATCGTGGCGCTTGTGTAATCGGCGGT 917
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
339 .....GlySerGlyPro.....ThrAlaProv 346

916 TCAATACCGAAGCGGAATACCGCGGTGCTGCTGCAACCAATTCGCC 867
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
346 alSerGlyGluThrArg...ThrSerValIleSerGlyThrAsnValPro 361

866 GCAGTAATTTGGGATACTTTCGACCCCAAAAGCGGTACGCAAGCGGTGG 817
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
362 Val.....SerGlyAlaProValThrProGlySerSerAlaGly 374

816 TTGTGTGACTTGAGAACCAACCAACCAATCACGCGCTCGGTGTTTCAGAC 767
| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
374 Y...SerSerGlyAlaProGlyThrGlyGlyProGlySerGluThrAla 390

766 GGCCTGTTGCAACCAACGTCGCGGTGCAATTCATCTTGATTAATGATG 717
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
390 exProLeuSerGlyAlaAlaGlyThrSerAlaThrGly..... 402

716 GTCCAAACGGTTTGTGTCACCGACCGGCTCAATGAAATGAATGCGGT 667
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

403 ...SerGlyThrSerIleProProSerGlyAla..... 412
666 GCCACTCAAAACCG.....GCCGATCGGGC 641
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
413 ProValThrProGluProProLeuIleSerThrGlyAlaSerAlaGlyP 429
640 CGCCGAATTCATGTGTTTCGATGTTGGCAGCATTTTCAGACGCGAGTCT 591
||||| ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
429 roProAlaSerSerGluSerThrVal.....Thr 438
590 GCGCCAGCTGCCTTACACACATGGATTTTACGCTCGGTCAACGGCTCAA 541
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
439 LeuProGlyAlaThrGlyThrAspValLeuArgSerGlyThrSerLeuPr 455
540 TACACAGCANACNTCGTCTGAAATCNCNGCGGCTCTTTTGATCACAACA 491
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
455 oValSerGlyGlyAlaVal...ThrProAlaProSerProGlyGlySerS 471
490 CAGGCTCTGCCGCGNAGCGGATGTTGTCATCGCATTTGACGAAGATGCGG 441
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
471 erAlaThrAlaGlyProGly...ValGlySerAlaThrThrValGlnAla 486
440 AACGGCTCGGCATCGACGCGGATTTTGTGTAACGCGGACGNTACGCGAG 391
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
487 SerGly...AlaThrGlyAlaAspValLeu.....ArgSe 497
390 CCAGCTCACAAACCG...GATTGGATCAGATTGCNCCNANNNTTTCGCG 344
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
497 rGlyThrSerLeuProValSerGlyValAlaValSerProGlySerSerP 514
343 CGCTTAAGTTTGGCAACGCTTCGCGCGGTAGCTTCGAACTCGATTTCG 294
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
514 ro.....GlyArgSerGlyAlaThrAlaValSerSerGlnGly 526
293 TGTTGCTCTCAACGCGCAATCACACCGACTGAAGTACGCTTTTCGCC 244
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
527 SerGlnProThrValAlaLeuSerGlyAlaThrGlyThrSerValGlyPr 543
243 G.....CGATGGATGGCGGCGATTTCCTGAAACNGCGCGGTAA 203
| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
543 oSerGlyThrArgPheSerSerSerAlaIleProAlaThrProGlySert 560
202 ACACG.....ACGCGCGGATNCTTTTCTCTTCA 174
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
560 hrThrGlyArgAlaAlaGlyAlaGlyThrProGlyValAspSerGln 576
173 AACGACCTTGGCCTTTTTCAGCGCATCGCTTCCTTTCAGCTTTTCATCNA 124
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
577 ThrAlaSerLeuProAlaAlaAlaArgProThrAlaLeuGly..... 590
123 GGGCGCATACCGGCATATCTTCGCCAAGCAACGCGACTTCG 81
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
591 .....ProGlyThrSerAlaProSerGlyGluThrSer 601

```

seq\_name: sp\_virus:O39781

seq\_documentation\_block:

```

ID O39781 PRELIMINARY; PRT; 866 AA.
AC O39781;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE MEMBRANE GLYCOPROTEIN.
OS Equine herpesvirus 1.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10326;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HH1;
RA Kirisawa R., Kobayashi T., Kawakami Y., Iwai H.;
RT "Nucleotide sequences of open reading frames 1, 24 and 71 of an
attenuated equine herpesvirus-1.";

```

RL J. Equine Sci 7:79-87(1996).  
DR EMBL; D88733; BAA20037.1; --  
SQ SEQUENCE 866 AA; 86463 MW

alignment_scores:		
Quality:	108.00	Length: 474
Ratio:	0.486	Gaps: 15
Percent Similarity:	46.835	Percent Identity: 20.464

alignment\_block:  
US-09-303-518D-127/rev x 039781

Align seq 1/1 to: 039781 from: 1 to: 866

1322 TCCAGCACCTTACGCCAACAGCGGGCCNTATTCGTATTTCGCCGGGCAGAC 1273  
 50 SerSerProThrThrSerProProThrThrSerSerSerProProThr 66  
 1272 GAAGCTGCACAAGCGAGGCTCTCTTCGTCCTCAATCCAAAGCAACCAATG 1223  
 66 rSerThrHisThrSerSerProSerSerThrSerThrGlnSerSerSert 83  
 1222 CTTGCGCGCTGTCGGTATCG...CCGACGATTAATTCGCCAAAGCAGG 1176  
 83 hrAlaAlaThrSerSerSerAlaProSerThrAlaSerSerThrThrSer 99  
 1175 GTAGGCAGGATGCTAGCGGCATTCAGCGCTCGCTAAGTCAACATCGGCAC 1126  
 100 IleProThrSerThrSerThrGluThrThrThrThrPro...Th 114  
 1125 CATGGCGCGTCCGCCACCGTTGACGGCTGTCGGGAACCTGAAGAGTTGT 1076  
 114 rAlaSerThrThrThrProThrThrThrAlaAla... 126  
 1075 TTTTCAGGAATGCCGAGGTCGTACGCGTGATGGAGTATTTCCTCGGC 1026  
 127 ...ProThrThrAlaAlaThrThrThrAlaValThrThr 138  
 1025 TGGCGGCGCAACCCAGCGAAGAGCTCTTTCGTCGGGCTTCTTCGATAAC 976  
 139 AlaAlaSerThrAlaAlaSerThrSerAlaGluThrThrThrAlaThrAl 155  
 975 GGAATCTGATTGGTGTAGCTCCCAATAATCTGCGCGCTGTGTGTAA 926  
 155 aThrAla...ThrSerThrProThrThr 164  
 925 TCGGCGGCTTCAATPACCGAACCGGAATC...ACG 894  
 164 hrThrProThrSerThrThrThrThrAlaThrThrThrValProThr 180  
 893 CGGTTGCTCGGTCAACCAATCCCGCGCAGTAATTCGATACATCTTCG 844  
 181 ThrAlaSerThrThrThrAspThrThrThrAlaAlaThrThrThrAlaAl 197  
 843 ACCCAAAACGGTACGCAAGAGCGGTGGTTGTGTGACTTTGAGAACCCACA 794  
 197 aThrThrThrAlaAlaThrThrThrAlaAlaThrThr...T 210  
 793 AAGCAATCACGCGCTCGGTGTTACAGCGGCTGTGTCAAACAAACGTCG 744  
 210 hrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAla 226  
 743 ATGCGCAATTACATCTTGATAATTGATGGTCCAAACGGTTTTGTTGCGAC 694  
 227 AlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThr 243  
 693 GACGCGCTCAATGAATCAATGTGCGTCCACTCAACCGCGCGGATGG 644  
 243 rThrSerSerAlaThrThrAlaAlaThrThrThrAlaAlaThrThrThrAla 260  
 643 GGCGCCGCAATTCAATGCTTTTCGATGTTGGCAGCAGCATTTTCAGACGC... 597

260 laAlaThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAla 276  
597 ..... 597

277 ThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrTh 293

596 .ACGCTGCGCCAGCTGCCCTACACACATGATTTACGCTCGCTCAAC 548  
||||:|||| :||:|||| |||  
293 rThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThr 310  
||||:|||| :||:|||| |||

547 GGCTCAATACACGACANACNTCGTCTGAATCTGAAATCNTCGNGCGCTTCTTGATC 498  
||||:|||| :||:|||| |||  
310 laAlaThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAla 326  
||||:|||| :||:|||| |||

497 ACAACACAGGCTCTCGGCGNAGCGGATGGTGCTCCATCGCATTCAGAA 448  
||||:|||| :||:|||| |||  
327 ThrThrThrGlySerProThrSerGlySerThrSerThrThrGlyAlaSe 343  
||||:|||| :||:|||| |||

447 GATGCGCAAGCGCTCGCATCGACGGCA...GGGATTTGCTGAACGAC 401  
rThrSerThrProSerAlaSerThrAlaThrSerAlaThrProThrSerT 360  
:||:|||| :||:|||| |||  
400 GGNATACGACGCGCAGTCCACAAACCGGATTGGATCAGATTGCNCCNNANT 351  
:||:|||| :||:|||| |||

360 hrSerThrSerAlaAla...AlaThrThr 368  
:||:|||| :||:|||| |||

350 TCNCTCGCGCGTTAGTTTGGCAAGCCTTCGGGGGGTAGCTTCAATC 301  
||||:|||| :||:|||| |||  
369 SerThrProThrProThrSerAlaAlaThrSerAla...GluSe 382  
:||:|||| :||:|||| |||

300 GATTTCGTGTGCTTCCACGGCAATCACGACCGACTGAATTCAGCGCT 251  
||||:|||| :||:|||| |||  
382 rThrThrGluAlaProThrSerThrProThrThrAsp...ThrThrThrP 398  
:||:|||| :||:|||| |||

250 TTTGCGCGCGATGGATGGCGGGATTTCCTGAAACNGCGCGCTAAAC 201  
||| ||||: ||||:|||| :||:|||| |||  
398 roSerGluAlaThrThrAlaThrThrSerProGluSerThrThrValSer 414  
||| ||||: ||||:|||| :||:|||| |||

200 .....ACACGCCCGGATGCTT 184  
||||: ||||: |||

415 AlaSerThrThrSerAlaThrThrThrAlaPheThrThrGluSerHisTh 431  
||||: ||||: |||

183 TTTGTCTCAACACACACTTGGCCCTTTTTCAGCGCATCGCTTCTCTTCA 134  
:||||:|||| :||||:|||| ||||:|||| |||

431 rSerProAspSerSerThrGlySerThrSerThrThrAlaGluProSerSerT 448  
||||: ||||: ||||:|||| |||

133 CTTTTCATCNAGGGCGCATACCGGATATCTTCGCCCAACGCGACT 84  
||||: ||||: ||||:|||| |||

448 hrPheThrLeu...ThrProSerThrAlaThrProSer...Thr 460  
||||: ||||: ||||:|||| |||

83 TCGGTAAATGACGGGCGCTCATAAATGACT.....TG 52  
||||: ||||: |||

461 AspGlnPheThrGlySerSerAlaSerThrGluSerAspSerThrAspSe 477

51 CTCGGCTCTGCCCGGATGGGC 30

477 rSerThrValProThrThrGly 484

seq name: sp vertebrate:073749

```
seq_documentation_block:
  id: 073749  PRELIMINARY:
    PRT: 1219 AA.
```

AC	073749;	
AD	01-AUG-1998 (TREMBLrel. 07, Created)	
AE	01-AUG-1998 (TREMBLrel. 07, Last sequence update)	
AF	01-AUG-1998 (TREMBLrel. 07, Last annotation update)	
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	NUCLEOPORIN NUP153 HOMOLOG (FRAGMENT).	
OS	Xenopus laevis (afrikan clawed frog).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipelidae; Pipidae;	
OC	Xenopodinae; Xenopus.	
OX	NCBI_TaxID=8355;	
FN	[1]	

RP SEQUENCE FROM N.A.  
 RX MEDLINE-98198465; PubMed-9531546;  
 RA Shah S., Tugendreich S., Forbes D.J.;  
 RT "Major Binding Sites for the Nuclear Import Receptor Are the Internal  
 Nucleoporin Nup153 and the Adjacent Nuclear Filament Protein Tpr.";  
 RL J. Cell Biol. 141:31-49(1998).  
 DR EMBL; AF045567; AAC41273.1; -;  
 DR InterPro; IPR001876; Znf-RanBP.  
 DR Pfam; PF00641; zf-RanBP; 5.  
 DR SMART; SM00547; Znf\_RB2; 5.  
 KW Porin.  
 FT NON-TER  
 SQ SEQUENCE 1219 AA; 123248 MW; 24EB9F96683696F6 CRC64;

alignment\_scores:  
 Quality: 108.00 Length: 516  
 Ratio: 0.448 Gaps: 28  
 Percent Similarity: 46.705 Percent identity: 21.124  
 alignment\_block:  
 US-09-303-518D-127/rev x 073749

Align seg 1/1 to: 073749 from: 1 to: 1219

1322 TCCAGCACCTTACGCAAGCGGGCCNPTATTCTGATTTCGCCGGG..... 1278  
 |||||::: ::::: ||| ||| |||||  
 312 SerSerAsnThrLysSerThrValProLeuSerSerThrProGlyLeuG1 328  
 1278 ..... 1278  
 328 yAspIlePheLysLysProAlaGlyMetTrpAspCysAspThrCysLeuV 345  
 1277 ..CAGACCAAGTCGACAAAGCGAGGTCTTCTGTCGTCGAATTCACACAA 1230  
 |||||::: ::::: ||| ||| |||||  
 345 aIcLnAsnLysAlaGluValThrLysCysValAlaCysGluThrProLys 361  
 1229 CCCATGCTTGC.....GCCGTGCGGTATCGCCGACGATTAAATC 1189  
 |||||::: ||| ||| ||| |||||  
 362 ProGlyThrCysIleLysAlaThrLeuLeuIleProSerThrThrLysSe 378  
 1188 GCGCAAGACGAGGTAGGACGAGGTCTAGCGGCATTCACCGCTCGTAAG 1139  
 |||||::: ||| ||| ||| |||||  
 378 rIleAsnProAlaThrAsnThrLeuAlaPheAlaSerCysSerAlaSerI 395  
 1138 TACCATCGGCACCATCGCGCGTCCGCCACCGTGTGACGGCTGCGTGAAC 1089  
 |||||::: ||| ||| ||| |||||  
 395 lEProAsnGluGluMetPheLysLysPro..... 404  
 1088 TTGAAGAGTTGTTTTCAGGAATGCCCCGAGGTGCTAGCGCTGATGGA 1039  
 |||||::: ||| ||| ||| |||||  
 405 .....MetGlySerTrpGluCysThrValCysHisMetG1 416  
 1038 GTATTTGTTC.....GGCTCGCGCGCAACCCACCGCA 1007  
 |||||::: ||| ||| ||| |||||  
 416 nAsnLysThrGluAspAsnThrCysValGlyCysLysAlaGluLysProG 433  
 1006 AC.....AGCTCTTTGCTG..... 993  
 |||||::: ||| ||| ||| |||||  
 433 lYThrValLysSerValProThrAlaAlaProSerGlyLeuLeuGlyLeu 449  
 992 .....CGGCTCTTCGATAACGGAATCTGATTGGTA 958  
 |||||::: ||| ||| ||| |||||  
 450 LeuHisGlnPheLysLysProThrGly.....SerTrp... 460  
 957 GCGTCCCAATATCGTGGCGGCTTGTGTAATCGCGGCTTCATACACG 908  
 |||||::: ||| ||| ||| |||||  
 461 .....AspCysAspValCysLeuIle.....GlnAsnL 470  
 907 AACCGGAATACGCGGTTGTCGCGTCAACCAATTCGCCCGCAGTAATT 858  
 |||||::: ||| ||| ||| |||||  
 470 yProGluAlaAlaLys.....CysIleAla 478

857 TCGCATACTTTCGACCCCAAAACGGTACCAAGAGCGGTGTTGTTGAC 808  
 |||||::: ||| ||| ||| |||||  
 479 CysGluSerAlaLysProGlyThrLysAlaGluProLysGlyThrPheAs 495  
 807 TTGAGAACACCAACCAACCAACCGTCCGTCGTTGTTACAGCGGCTGTG 758  
 |||||::: ||| ||| ||| |||||  
 495 p.ThrValLysAsnSerValSerValAlaProLeuSerSerGlyGlnLeu 511  
 757 CAACAAACGTCGATGCAATATACATCTTGATAATTGATGTCACAAAG 708  
 |||||::: ||| ||| ||| |||||  
 512 GlyLeuLeuAspGlnPheLysLysSerAlaGlySer...TrpAspCysAs 527  
 707 GTTTGTTGTTG.....CACCGACCGGCTCAATGAATGAATGTCGTCGCC 664  
 |||||::: ||| ||| ||| |||||  
 527 pValCysLeuValGluAsnLysProGluAlaThrLysCysValAlaCysG 544  
 663 ACTCAAAACGCGCGGATCGGGCGCGCAATCATGTTTCGATGTGG 614  
 |||||::: ||| ||| ||| |||||  
 544 luThrSerLysProGlyThr.LysAlaGluLeuLysGlyPheGlyThr.S 560  
 613 CAGCATTTTCAGACGCGACGTCGCGCAGCTGCTTACACACATGGATT 564  
 |||||::: ||| ||| ||| |||||  
 560 exThrPheSerSerGlyThrAlaAlaProThr..... 570  
 563 TTACGCTCGGTCAACACGCTCAATACAGCANACNTGTCGTGAATCN... 516  
 |||||::: ||| ||| ||| |||||  
 571 PheLysPheGlyValGlnSerSerAspSerThrAlaGluLeuLysSerG1 587  
 515 ....TCGNCGGCTTCTTGTATCAACACACAGGCTGTCGCCNAGCGGAT 470  
 |||||::: ||| ||| ||| |||||  
 587 yAlaSerThrSerGlyPheAlaLysSerIleGlyAspPheLysPheGlyL 604  
 469 TGTGTCATCGCATGACGAAGATGCGAACGCGTCGCGATCG..... 426  
 |||||::: ||| ||| ||| |||||  
 604 euValSerAlaSerThrThrThrGluGluThrGlyLysLysSerPheThr 620  
 425 .....ACGCGAGGGATTTTGTCTGAACGG 403  
 |||||::: ||| ||| ||| |||||  
 621 PheGlySerSerThrThrAsnGluValSerAlaGlyPheLysPhe...G1 636  
 402 ACGNTACGACGCGAGTCACCAACCGGATTCGATCAGATTCGNCNNA 353  
 |||||::: ||| ||| ||| |||||  
 636 yIleAlaGlySerAlaGlnThrLysProAspThrLeuSer...GlnSerI 652  
 352 NTTTCGCGCTTAAGTTTTCGCAACGCTTCGGCGGCTAGCGTTCGAAC 303  
 |||||::: ||| ||| ||| |||||  
 652 hrThrSerGlyPheThrPheGlySerVal.....SerAsn 663  
 302 TCGATTTTCGTCGTCCTTCAACGCGCAATCACGACCGACTGAGTACGCG 253  
 |||||::: ||| ||| ||| |||||  
 664 ThrValSerLeuAlaPro.....AlaAlaThrSerSerSerSerThrG1 678  
 252 CTTTTCGCGCGATGCGCGCGATTTTTCCTGAAACNGCGCGGTAA 203  
 |||||::: ||| ||| ||| |||||  
 678 yLeuGln.....ValAlaAlaAlaIleAlaAspSerAsnLeuAla 692  
 202 ACAC..... 198  
 |||||::: ||| ||| ||| |||||  
 692 hrThrAlaAlaLeuLysSerAlaGluLysLysAlaGluAlaProThr 708  
 197 ...ACGCGCGATNCTTTTGTCTTCAAAACAGCACTTGGCCTTTTGTGAC 151  
 |||||::: ||| ||| ||| |||||  
 709 lIleThrProPheSerPheGlyLysThrAspGlnAsn.....Ly 721  
 150 GGCATCGCTTCTTCTGACTTTCATCAGGGCGCATACCGCATATCTT 101  
 |||||::: ||| ||| ||| |||||  
 721 sGluThrAlaSerThrSerPheValPheGlyLys.....L 733  
 100 CGCAACCAACGCACTTCGTTAATGACGGCGCGCTCATTAATGACTTGC 51  
 |||||::: ||| ||| ||| |||||  
 733 ysAspGluLysThrAspSerAlaProThrGlySerSer.....PheAla 747  
 50 TCCGGTCTGCCGCGGATGGC...AGTTTAGACCTTTTGTG 12

```

seq_name: sp_human:Q9GZV6
seq_documentation_block:
ID Q9GZV6 PRELIMINARY; PRT: 1187 AA.
AC Q9GZV6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE MUC4 PROTEIN SPLICED VARIANT SV14 (MUC4 PROTEIN VARIANT VII).
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREATIC TUMOUR;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=20381033; PubMed=10920259;
RA Choudhury A., Moniaux N., Winpenny J.P., Hollingsworth M.A.,
RA Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin cDNA and its variants in pancreatic carcinoma.";
RL J. Biochem. 128:233-243(2000).
DR EMBL; AJ400851; CAC14136.1; -
DR EMBL; AJ400633; CAC10062.1; -
SQ SEQUENCE 1187 AA; 121925 MW; 62E4C93F86443B8F CRC64;

alignment_scores:
Quality: 107.50 Length: 446
Ratio: 0.578 Gaps: 17
Percent Similarity: 41.704 Percent Identity: 21.076

alignment_block:
US-09-303-518d-127/rev x Q9GZV6 ..
Align seg 1/1 to: Q9GZV6 from: 1 to: 1187
1340 CCTTCTCTTCAAGGTTTCCAGCACCTTACGCAACAGCGGCCNTATTC 1291
||||| ||| :::::|||||
810 ProSerSerSerGlyAlaSerGlyThrProSerGlySerGluGlyI 826
1290 GTATTTCGCCGCGCAGCAGCAAGTGCACAAACGCGAGTCTTCTTCGTCCA 1241
||||| ||| :::::|||||
826 eSerThrSerGlyGluThr.....ThrArgPheSerSerAsn 839
1240 ATTCCAAGCAACCAATGCTGGCGCTCGGTATCGCGCAGATTAA 1191
||||| ||| :::::|||||
839 roSerArgAspSerHisThrThrGlnSerThrGluLeuSerAla 855
1190 TCCGCAAAAGCAGGCTAGGAGGATCTAGCGGATTCAGCGGTCTGTA 1141
||||| ||| :::::|||||
856 SerAlaSerHisGlyAlaIleProValSerThrGlyMetAlaSerSeril 872
1140 AGTACCAATCGGCACCATCGCGGTCGCCACCGTGTGCGGTGCGTGA 1091
||||| ||| :::::|||||
872 eValPro...GlyThrPhe.....HisProThrLeuSerGluAla... 884
1090 ACTTGAGAGTTGTTTTCAGGAATATGCCAGGTCCTACGCGGATG 1041
884 .....
1040 GAGTATTGTCGGCTCGCGCGCAACCCAGCACGCTCTTGTCTGGC 991
||||| ||| :::::|||||
885 .....SerThrAlaGlyArgProThrGlyGlnSerSerProThrSe 998

```

```

990 GCCTTCTTCGATACGGAATCTGATTGTGTAGCTCCCAATAATCGT 941
||||| :::::
898 rProSerAla..... 901
940 GCGGCGCTTGTAAATCGCGCGCTTCAATACGACCGGAAATACACGGG 891
::: ||| ::::: |||
902 ..SerProGlnGluThrAlaIleSerArgMetAlaGlnThrGlnArg 917
::: |||
890 TTGCTGTGCTCAACCAATTCGCCCGCAGTA.....ATTG 856
::: ||| ::::: |||
918 ThrGlyThrSerArgGlySerAspThrIleSerLeuAlaSerGlnAlaTh 934
::: |||
855 CGATACTTTCGCA.....CCAAAACGGTACGCAAGAGCGGTGTTGT 812
||||| ::::: |||
934 rAspThrPheSerThrValProThr..... 943
||| ::::: |||
811 TGACTTGAACACCCCAAGCAATACGCGCTCGTGTTCAGACGGCT 762
||| ::::: |||
944 .....ProSerIleThrSerSerGlyLeuThrSerPro 955
||| ::::: |||
761 GTTGCAACAAACGTCGATGCGCAATTACATCTTGATAATTGATGTC 712
||||| ::::: |||
958 nThrHisThrLeuSerProSerGlySerGlyLys.....GlnThr 958
::: |||
661 TCAACACGCGCGGATCGCGCGCCGCAATTCATGTTTCGATGTGGCA 612
||| ::::: |||
970 .....ThrPheThrAla 974
611 GCATTTTCAGACGCGCTGCGCGCAGCTGCCTTACACACATGGATTT 562
::: ||| ::::: |||
975 LeuIleSerAsnAlaThrProLeuProVal..... 984
561 ACCTCGTCAACGCGCTCAATACGACGACNACNTCTGCTGAATCNTGN 512
||||| ::::: |||
985 .....ThrSerThrSerSerAlaSerThrGlyH 994
511 CGGCTTCT...TTGATCAACACACAGGCTGTCGCCGACGCGGATGGT 465
||||| ||| ::::: |||
994 isAlaThrProLeuAlaValSerSerAlaThrSerAlaSerThrValSer 1010
464 TCCATCCATTCAGAGATGCGCAACGCGCTCGCATCGACGCGAGGAT 415
||| ::: ||||| ::::: |||
1011 SerAspSerProLeuLysMetGluThrSerGlyMetThrThrProSerIle 1027
::: ||| ::::: |||
414 TTTGCTCAACGCGGNTACGCGCAGCTCCACAAACCGGATGGATCA 365
::: ||| ||||| ::::: |||
1027 uLysThrAspGlyArgArgThrAlaThrSerProPro..... 1041
364 GATTGCNNCANNNTCTTCGCCGCTTAAGTTGCCAACGCTTCGGCGCG 315
::: ||| ::::: |||
1042 .....ThrThrSerGlnThrIle..... 1047
314 TAGCGTTTCAACTCGATTTCGCTTCGCTTCAACGCAATACGACCGGA 265
||||| ::::: |||
1048 .....IleSerThrIleProSerThrAlaMetHisThr... 1058
264 CTGAAGTACGCGCTTTTCGCCGCGGATGGATGGCGG.....ATT 224
||| ::::: |||
1059 .....ArgSerThrAlaAlaProIleProIleL 1068
223 TGCTGTAACNCGC.....GCGTAAAC 201
||||| |||
1068 euProGluArgGlyAlaIleGlnIleGluPheProSerSerProMetGly 1084
200 ACCAGCGCGGATCTTTTCTCTTCAACAGCACCTTGGCGCTTTTTCAC 151
||||| ::::: |||
1085 GlnThrProGlyThrThrProSerSerSerGlyGlyProThrProThr 1101
150 GGCATCGCTTCTTCTGACTTTTCATCNAGGCGCATACCGGCATATCTT 101

```

```

1101 oHisSerSerArgArgLeuAlaSerProLeuAlaProLeuSerValI 1118
100 CGCAACGACGCGACTTCGGTAATACGGGCGCGTCA 63
1118 leProSerThrSerGlnThrMetAlaArgSerSer 1130
seq_name: sp_human:Q14395

```

seq\_documentation\_block:

```

ID Q14395 PRELIMINARY; PRT; 505 AA.
AC Q14395;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MUCIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=STOMACH;
RX MEDLINE=93194895; PubMed=7680650;
RA Toribara N.W., Robertson A.W., Ho S.B., Kuo W.L., Gum E., Hicks J.W.,
RA Gum J.R. Jr., Byrd J.C., Siddiki B., Kim Y.S.;
RT "Human gastric mucin. Identification of a unique species by expression
RT cloning.";
RL J. Biol. Chem. 268:5879-5885(1993).
FR EMBL; L07518; AAB61945.1; -.
FT NON_TER 1 505
FT NON_TER 505 505
SQ SEQUENCE 505 AA; 51967 MW; 45D7CB1E6050E07D CRG64;

```

alignment\_scores:

```

Quality: 107.00 Length: 310
Ratio: 0.690 Gaps: 14
Percent Similarity: 50.000 Percent Identity: 22.258

```

alignment\_block:

US-09-303-518d-127/rev x Q14395 ..

Align seg 1/1 to: Q14395 from: 1 to: 505

```

926 ATCCGCGCGTTCATACGACCGGAATACGCGGTGTCTCGGTCAAC 877
87 ValAlaProThrSerThr...ThrThrIleThrProAsnProThrSerTh 102
876 CAATTCCGCCCGCAGTAATTTCCGATACTTTCGACACCAACAGGTACGCA 827
102 rArgThrArgThrProValAlaHisThrAsnSerAlaThrSerSerArgp 119
826 AGAGCGGTGTTGTTGACTGTGAGAACCCCAAGCAATCAGCGCTCG 777
119 roProProPheThrThrHisSerProThrGlySerSer..... 133
776 GTGTTACAGCGCTGTGTGCAACACGTCGATGCCAATTCATCTTG 727
134 .....PropheSerThrGlyProMetThrAlaThrSer.. 145
726 ATAATTGATGTCACCAACGGTTTGTTCACCGCGGCTCAATCAAT 677
146 .....PhelystThrThrThrThrThrThrThrThrThrThrThr 157
676 GAATGCGGTCCCACTCAACCGCGGATGCGGCGCGGAATTCATGT 627
158 .....ProGlnThrThrPro 162
626 GTTCGATGTTGCGAGCATTTTCAGACGCGACGCTCGCGCCGCGCTT 577
163 LeuThrHisValProProPheSerThrSerLeuValThrProIleThr.. 178

```

```

576 ACACACATGATTTTACGTCGCTCAACGCGCTCAATACCAACANCA... 531
179 .HisThrValIleThrProThrHisProGlnMetSerThrSerAlaTyri 195
530 .....CNTGCTCTGAAATCN 516
195 leHisSerThrProThrGlyThrIleAlaSerProThrThrValLysAla 211
515 TCGNCGGCTCTTGTATCACACACAGGCTCGCCGACGCGGATGTTGTT 466
212 ThrArgSerThrTyThrAlaProLeuMetThrAlaThrThrSerArgil 228
465 GTCCATCGCATTCAGCAAGATGCGACGCGCTCGGCATCGACGCGAGGA 416
228 eSerGlnAlaHisSerSerIleSerThrAlaLysThrSerThrSer.... 243
415 TTTTGTGTAACGCGGNTACGCGCAGTCCCAACACCGGATTTGGATC 366
244 .....LeuHisSerHisAla...SerSerThrHisHisProGlu..... 255
365 AGATTCCNNNNANTTCNTCGCGCTTAAGTTTGCCAAACGCTCGGGCGC 316
256 .....ValThrProThrSerThrThrAsnValThrProLy 267
315 GTAGCGTTCCG.....AACTCGA 299
267 sSerThrSerArgAspThrSerThrProValThrHisThrThrSerAlat 284
298 TTTTCGTGTTGCTTCAACGCGCAATCAGACCGACTGAAGT...ACGCGC 252
284 hrSerSerArgProProThrProIleThrThrHisSerSerProThrArg 300
251 TTTTCCGCGCATGATGCGCGGATTTGCTGAAACNGCGCGGTAA 202
301 SerSerPro.....LeuSerSerThrGlyProMetTh 311
201 CACACGCGCGGATNCITTTTGTCTTCAACACGACTTGGCCTTTTGTGA 152
311 rAlaThrSer.....IleLysThrThrThrThrTyPro.....T 323
151 CGGATCGCCTTCCTTCATCATCAGGCGCATACCGCATATTTCT 102
323 hrProSerHisProGlnThrThrLeuThrThrHisValProProPheSer 339
101 TCGCAACGACGCGACTTCGGTAATGACG 72
340 ThrSerSerValThrProSerThrHisThr 349

```

seq\_name: sp\_invertebrate:Q9V515

seq\_documentation\_block:

```

ID Q9V515 PRELIMINARY; PRT; 746 AA.
AC Q9V515;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CG8181 PROTEIN.
DE NCBI_TaxID=7227;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,

```

```

RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AF03835; AAF59007.1;
DR FlyBase; FBgn0033361; CG8181.
DR InterPro; IPR002965; P_Rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 746 AA; 78593 MW; FB6F9F8DA3027334 CRC64;

alignment_scores:
  Quality: 107.00      Length: 524
  Ratio: 0.476        Gaps: 21
Percent Similarity: 42.939 Percent Identity: 20.229

alignment_block:
US-09-303-518D-127/rev x Q9V515
..

Align seg 1/1 to: Q9V515 from: 1 to: 746

1334 TTCTCAANGGTTTCCAGCACCTTACGCAACAGCGGCCNTATTTCGTATT 1285
||||| |||:||||| |||:||||| |||:|||||
53 PheSerHisGluSerThrLeuProThrSerAlaLeuPhelys....Le 68

1284 GCCCGGCAGACGAAGCTGCACAAACGAGGCTCTCTTCGTCCATTC. 1236
| |||:||||| |||:||||| |||:||||| |||:|||||
68 uThrThrGluSerSerLeuSerSerSerSerSerSerGlySerSerThra 85

1235 .....AAGCAACCCCAATGCTTCGCGC 1215
||||| |||:||||| |||:||||| |||:|||||
85 laHisHisHisThrThrAsnGluLeuLeuLysAsnValThrGlnSerTy 101

1214 CTGTCGGTATCCGCGAGATTAAATCCGCAAGAGAGGTTAGGAGGAT 1165
|||||:||||| |||:||||| |||:||||| |||:|||||
102 LeuSerThrAlaLysProAlaProValArgThrThrSerAsnGlyThrLe 118

1164 GTCTAGCGGCATTCAGCGCTCGTAAGTACCAATCGCACCATCGCGCGT 1115
:||||| |||:||||| |||:||||| |||:||||| |||:|||||
118 uSerThr.....ThrArgA 123

1114 CGCCACCGTTACGGCTGTGCTGAACCTTGAGAGTTGTTTTCAGGAAA 1065
|||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
123 rgProProLeuAlaHisThrThrHisLysAsnSerHisTyr..... 136

```

```

1064 TGGCCGAGGTCGTACGCGTGATGAGTATTGTCCGCTGGCGGCAAC 1015
|||||:||||| |||:||||| |||:||||| |||:|||||
137 .....ValValArgThrThrMetLysProSerAlaAlaProThrTh 150
1014 CCAGCGGAACACTCTTTCGCTGGCGCTCTTCGATAACGGAATCTGAT 965
| |||:||||| |||:||||| |||:||||| |||:|||||
150 rThrileThrProThrValLysProProArgArgThrThr..... 165
964 TGTGTGAGCGTCCCAATAATCGTCGGCGCTTGTGTA..... 927
166 .....AlaLysProValLeuThrThrAlaProileValGluGlnArgile 180
927 ..... 927
181 GluThrThrSerValProThrLysPheValThrPheGlnIleValGluTh 197
926 .ATCGCGCGCTTCAATACCGAACCG..... 903
197 rThrThrProAlaValThrGluProThrHisGlnGlnGluThrThrTh 214
902 ..... 900
214 hrThrThrArgArgHisThrArgProThrGlySerSerSerValAlaSer 230
899 ATCAGCGGTTGTCTCGTCAACCAATTCGCCGCGAGTAATTTCGATAC 850
||| |||:||||| |||:||||| |||:||||| |||:|||||
231 ArgThrThrSerLysThrThrThrThrThrThrThrThrThrThrTh 247
849 TTTCGCAACCAACAGTACGCAAGAGGCTGTTTGTGACTTGAGAA. 801
| ||||| |||:||||| |||:||||| |||:||||| |||:|||||
247 rAlaAlaProProThrThrThrSerThrArgAlaAlaThrThrThrGln 264
800 .....CCACCCAAAGCAATCACGCGTCG..... 777
||||| |||:||||| |||:||||| |||:||||| |||:|||||
264 rgProThrProProLeuArgThrThrThrSerLysProHisLysThrHis 280
776 GTGTTACAGCGCTGTTCACCAACAGCGCTGCAATTCATCTTG 727
||||| |||:||||| |||:||||| |||:||||| |||:|||||
281 SerSerArgArgProAlaProAlaLysLysProValSerThrThr..... 295
726 ATAATTGATGTCACAAACGGTTTGTTCACCGCGCTCAATGAAT 677
||||| |||:||||| |||:||||| |||:||||| |||:|||||
296 .....ThrThrThrAlaSerGlyThrGlnSerThrArg. 306
676 GAATGTCGTCGCTCAACACCGCGGATCGCGCGCGCAATTCATGT 627
||||| |||:||||| |||:||||| |||:||||| |||:|||||
307 .....LysProAla.....AspSer 311
626 GTTTCGATGTCGAGCATTTTCAGACGCGACGCTCTGCG.....CC 586
||||| |||:||||| |||:||||| |||:||||| |||:|||||
312 LeuAlaSerGlySerThrThrAlaAsnGlyThrThrAlaThrThrLeuPr 328
585 AGCTGCCTTACACACATGGATTTTACGCTGGTCAACAGCGCTCAATACCA 536
||||| |||:||||| |||:||||| |||:||||| |||:|||||
328 cAlaArg.....ArgProValThrGlnThrSerSerS 339
535 GCANACNTGCTGAAATNTCGNCGGCTCTTGTGATCAACAAC..... 492
339 erSerThrThrThrThrSerThrThrThrThrThrThrThrValphe 355
492 ..... 492
356 LysAspGluLeuProLysAsnArgThrThrArgLeuProGluArgTh 372
491 .....ACAGGG.....TCTG 482
372 rThrGlnAlaProArgProArgProLysProThrGlyGluLeuValLysV 389
481 CGCGNAGCGGATTGTCCTCATCGCATTCAGCAAGATGGCAACGGCTCG 432
|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
389 alProAlaLeuLeuAlaGluProThrThrThrThrThrThrThrLysGlySer 405
431 GCATCGACGCGGAGGATTGCTGAACGCGGNTACCGACGCGAGTCCA 382

```



```

:::||||: SerSerSer 413
381 CAAACGGATGGATCAGATGCGNCCNNTTCNTCCGCGTT ..... 339
||||| SerSerSer 413
413 oLysProPro ..... AlaSerSerThrThrProLeuValThrL 426
338 ..AAGTTTGCACAGCTGGCGCGCTAGCGTTCGAACCTCGATTTCGTCG 291
||| ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 442
426 yLysLysProThrThrThrThrThrThrThrThrThrThrThrThrThr 241
290 TTCCCTTCACGCAATCAGCAGCAGCTGAAGTACGCGCTTTTCGCGCG 443
443 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 459
240 ATGGATGGCGCGGATTTCCTTGAACGCGCGGTAAACACACGCGCG 191
459 o ..... ProThrThrThrThrThrThrThrThrThrThrThr 470
190 GATNCTTTTCTTCAACAGCAGCTTGGCGCTTTTTCGCGCAGCGCT 141
470 leThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 486
140 TCCTTGACTTTCATONAGGCGCATACGCGCATATTCTTCGCCAAGCAA 91
487 LeuLeuThr ..... ThrGluProProThrSerAlaProLeuTh 499
90 CGGACTTCGTAATGACGGGC 69
499 rThrThrThrLysLysThrGly 506
seq_name: sp_human:Q9Y2N2

seq_documentation_block:
ID Q9Y2N2 PRELIMINARY; PRT; 1294 AA.
AC Q9Y2N2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE WUGSC:H_D0953A04.3A PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99063792; PubMed=9847074;
RA Sulston J.E., Waterston R.;
RT "Toward a complete human genome sequence.";
RL Genome Res. 8:1097-1108(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Cordes M., Bauer C., Holmes A.;
RT "The sequence of Homo sapiens PAC clone RP5-953A4.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006014; AAD28063.1;
DR InterPro; IPR002965; P_rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 1294 AA; 131942 MW; 6FC9761CC44162C2 CRC64;

```

```

alignment_scores:
  Quality: 107.00
  Ratio: 0.546
  Percent Similarity: 46.118
  Gaps: 17
  Percent Identity: 22.824

```

```

alignment_block:

```

```

US-09-303-518d-127/rev x Q9Y2N2

```

```

Align seg 1/1 to: Q9Y2N2 from: 1 to: 1294

```

```

1115 TCGCCACCGTTGACGGTGTGCTGAAC .....TTGAAGAG 1081
||||| SerProProThrSerLeuLeuAlaProSerThrAsnProLeuLeuGluSe 577
1080 TTTGTTTTTTCAGGAATAAGCGGAGGCTGTACGCGTGTATGAGTATTGT 1031
||||| rLeuLysLysMetGlnThrProProSerLeuProProCysProGluSerA 594
1030 CCGGCTGGCGGCAACCCAGCCGACGAGC .....TCTTTG 996
||||| laGlyAlaAlaThrThrGluAlaLeuSerProProLysThrProSerLeu 610
995 CTGCGGCTTCTTCGATAACGGAATCTGATTGTGTAGCTGCCCAATA 946
||||| LeuProProLeuGlyLeuSerGln ..... 618
945 ATCGTGGCGGCTTGTGTAATCGCGCG .....TTCAATACCGAAC 905
619 ....SerGlyProProGlyLeuLeuProSerProSerPheAspSerLysP 634
904 CGGAATACCGCGGTTGTCT .....GCGTCAACCAATTCGCCCGCA 864
634 roProThrThrLeuLeuGlyLeuLeuProAlaProSerMetValProAla 650
863 GTATTTTCGATACCTTTCGACCCCAACAGGTACGCAAGAGGCGGTGT 814
||||| Thr .....AspThrLysAlaProProThrLeuGlnAlaGluThrAlaTh 665
813 GTTGACTTGAGAACCA .....CCCAAGCAATCACGCGCT 779
665 r .....LysProGlnAlaThrSerAlaProSerProAlaProLysG 679
778 CGGTGTTTCAGCGGCTGTTTGCACCAACAGCTCCGATG ..... 741
679 InSerPheLeuPheGlyThrGlnAsnThrSerProSerProAlaAla 695
740 ...GCAATTACATCTGATAATTGATGTCGTCACCAACGGTTTGTTCGACC 694
696 ProAlaAlaSerSerAlaSerProMetPheLysProIlePheThrAlaPr 712
693 GACCGGCTCAATGAAATGATGTCGTCGTCACCAACCAACCG ..... 654
712 oProLysSerGluLysGluGlyProThrProGlyProSerValThra 729
654 ..... 654
729 laThrAlaProSerSerSerSerLeuProThrThrThrThrAla 745
653 .....GCCGGATGGCGGCGCGCAATTCATGTGT 625
746 ProThrPheGlnProValPheSerSerMetGlyProProAlaSer...Va 761
624 TTCGATGTTGGCAGCATTTTCAGACGCGCAGTCTGCGCAGCTGCGCTTAC 575
761 lProLeuProAlaProPhePheLysGlnThrThrThrProAlaThrAlap 778
574 ACACATGGATTTCACCTCGCTCAACCGCTCAATACGACGACNACNTCOT 525
778 roThr ..... 779
524 CTGAAATCTCGNCGGCTTCTTTGATCACACACAGGCTGTCGCCGNAG 475
780 .....ThrThrAlaProLeuPheThrGlyLeuAlaSerAlaThrSe 793
474 CGGATGTTGTCATCGCATTCAGCAAGATGCGCAACGCTCGGCATCG. 426
793 rAlaValAlaProIleThrSerAlaSerProSerThrAspSerAlaSerL 810
425 .....ACGGCAGGATTTTGTCTGACGCGGACGNTAGCGACGCGCATCCAC 381
810 ysProAlaPheGlyPheGlyIleAsnSerValSerSerSerVal... 825

```



```

1410 .GlnProHisSerHisPheProGlnSerThrGlnMetLeuProGlnSerG 1426
887 .....
1426 lyAsnPheSerSerValSerHisLeuLeuThrThrHisProMetSerSer 1442
878 ACCAATTGCGCGCCAGTAATTTGCGATCTTTGCGCACCCCAAAACGGTAGC 829
1443 GlnAsnGlnProMetValArgCysGlySer.....ThrLeuTyf 1455
828 CAAGACGGGTGGTTTGTGACHTGAGAACCCCAAGCAATCAGCGCT 779
1455 rSerGlnSerSerAlaAlaThrAlaProProSerAlaAlaAla.... 1470
778 CGGTGTTTCAGCGCGCTGTGCAACCAACGTCGCGATGCAATTCAT 1479
1471 .....AlaValSerAsnPheThrProSerVal..... 1479
728 TGATAATTGATGTCACAAACGGTTTGTGTCACGCGCGCTCAATGAA 679
1480 .....LeuAlaValGlnSerLeuThrThrAlaValThrSerSer..... 1492
678 ATCAATGTCGTCGCCACTCAACCGCGGATGCGGCGCGCGAATTCAT 629
1492 .....
628 GTGTTTCGATGTTGCGCAGCATTTTCAGACGGCAGCTCTGCGCGCAGTCC 579
1493 .....SerSerSerProSerThr 1498
578 TTACACACATGATTTAGCTCGTGC.....AAACG 547
1499 LeuSerSerSerValIleGlnValIleSerProLysGlyGluSerPr 4515
546 GCTCAATACCAACANCTCGTCTGAATCTGCGCGCTTCTTTGATCA 497
1515 cCysAsnLysAspArgAspSerSerTyrSerSerProAlaAsnAlaValV 1532
496 CAACCAACA.....GGTCTGCGC.... 480
1532 aLThrCysAlaProThrThrProIleValSerSerGlySerAlaArg 1548
479 .....GCNACGGATTGGTGC 463
1549 ProThrProProLeuSerAsnCysThrSerMetGlyIleGlyMetValAs 1565
462 CATCGCATTCAGAGATGCGCAACGGCTCGCA..... 429
1565 nAlaAlaSerThrAlaArgSerSerCysAsnAlaIleSerProLeuSerI 1582
428 .....TCGACGGCAGGGATTTGCTGAACGGCAGGNTACGCGCGCAGTC 384
1582 leProAlaThrAlaGlyIleHisValSerAlaThr..... 1593
383 CACAACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 334
1594 ...AsnProSerPheGlnSerSerTyrPheProThrProLeuAlaPr 1609
333 TGCCAAACGCTTCGGCGCGTAGCTTCCAACTCGATTCGTCGTGCTT 284
1609 oProProSerSerProSerProAlaThrSerSerAlaAlaIleIleSerS 1626
283 CAACGGCATTCAGCACCGCTGAAAGTACGCGCTTTTCGCGC..... 243
1626 erSerAla.....SerGlnPheAsnProAlaValSer 1636
242 CGATGATGCGCGGATTTTCCTGAACNGCGCGCTGAACACACCGCC 193
1637 HisSerMetSerSerIleValThrThrAlaGlyAlaThrThrThrAl 1653
192 CGGATNCTTTTGTCTCAACACACCACTTGGCCTTTTTCACGGCATCGC 143

```

```

1653 aserSer.....ValThrGlnP 1659
142 CTTCCTTGACTTTCATCAGGGCGCATACCGCATATTTCTTCCCAAGC 93
1659 roSerValAlaAlaIleSerAsnProValThrAsnThrProHisProphe 1675
92 AACCGGACTTCGGTA 78
1676 SerAlaGluSerLeu 1680
seq_name: sp_bacteriap:Q9I3F4
seq_documentation_block:
ID Q9I3F4 PRELIMINARY; PRT; 352 AA.
AC Q9I3F4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE HYPOTHETICAL PROTEIN PA1563.
GN PA1563.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Hickey C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Garber R.L., Golltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.L., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AE004584; AAG04952.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 352 AA; 40156 MW; 3030DSABF0513EBB CRC64;

```

```

alignment_scores:
Quality: 106.50 Length: 302
Ratio: 0.783 Gaps: 13
Percent Similarity: 45.033 Percent Identity: 22.848
alignment_block:
US-09-303-518D-127 x Q9I3F4
Align seg 1/1 to: Q9I3F4 from: 1 to: 352
601 TCTGAAATGCTCCCAACATCGAACACATGAAATTCGGCGCCGCATCC 650
21 AlaGluHisAlaAlaThrLeuGluIleProGlyTyrAlaLysSerLysPr 37
651 GGCCGGTTCAGTCGCGCAGCAGCATTCATTTCATT.....GAGCCGGTCG 694
37 oAla.....SerAlaHisValGluPheValCysGlnAspAlaAspG 51
695 GTCAACAAACACCGTTTGGACCATCAATATCAAGATGTAATT..... 738
51 lyAlaGluArgLeuMetArgArgLeuArgPheAlaAspLeuIlePhePro 67
739 .....GC 740
68 ArgGlnTrpAlaArgGlyProGlyPheIleGluLeuProGluSerGlnAr 84
741 CATCGGACGTTTGTTCGAACAGCGCGTCTGAACACCGAGCGCGTATG 790
84 gIle..GluValLeuLeuAlaGluLeuAlaSerTyrProValCysGlySer 100
791 CTTTGGGTGGTTCTCAAGTCAACAAACACCGCTCTTGGGTACCGTTTG 840

```



```

228 gHisSerGlyAlaHisLeuGlyGlyGluLeuAspArgValAlaGlyVala 245
575 GTAAGGCGAGTGGCGAGAGCTGCCCTCAAAATCTGCCACATCGAA 524
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
245 laLeuValAlaGlnAlaProValAlaGluGlu 255
625 ACACATGAATTCGGCGCCCGCATCCGCGGCTTTGAGTGGCAGC...CA 671
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
256 ..... 256
672 CATTCAT...TTCATTGACCGGTGGTGCACAAACACCGTTGGACCA 718
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
263 svaHisValValGluProAlaGlyArgGluAspAlaProAla 280
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
719 TCAATTATCAAGATGAATTCATCCGAGCTTTGTTTTCACAGGCGGT 768
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
280 rglyHisGlyAspProAlaAlaValAlaLeu.....Glu 291
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
769 CTGAACACCGAGCGGTGATTCCTTTGGGTGTTCTCAAGTCAACAAACC 818
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
292 HisArgAlaAspHisProValAlaLeu...GlyAspGlnLeuAspGlnAr 307
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
819 ACGCCCTTG.....CGTACCGTTTGGGTG 844
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
307 gGlyLeuAlaProAspArgAlaGlyAlaGlnArgAlaValGluGlnP 324
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
845 CGAAAGTATCGCAATTTACTGCGGGCGAATGTTGACGACGACCAACGC 894
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
324 roGlyArgGluArgProAlaGlyGlnValAlaAlaAspHisHis 340
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
895 GTGATTTCGGTTCGGTATTGAACGCGCGATTACACAAAGCGCGCACGA 944
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
341 .....AlaAlaAspAlaLeuSerGlyAlaAlaGlnAspAlaArgGlnAl 355
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
945 TTATTGGGACCTACCAACATCATTC..... 974
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
355 aLeuAlaGlyLeuAlaArgAspGlnValHisProLeuValValArgThrG 372
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
975 .....CGTTATCGA.....AGAACGCCGACGCAAGA 1001
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
372 lyAspArgHisArgAspArgCysLeuAspAlaArgProGlnGlnArg 388
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1002 GCTGTTCGGCTGGTGGCGCGACCGGACAAATATCTCCATCAGCGTA 1051
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
389 Pro.....GlyLeuAlaGlnHisArgArgValGluArgLeuAlaLe 402
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1052 CGAC.....CTCGCGCATTCCTGAAAAACAACTCTTCAAGTT..... 1091
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
402 uAspAlaAlaProArgGlyValAlaAlaArgGlnLeuArgValValG 419
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1092 .....CACGACGCGCTCAACGGTGGCGACCGGCAT 1124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
419 lyValAlaAlaArgProHisGlnLeuGluArgArgProGlnHis 435
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1125 .....GTGTC.....GATTGGTACTTACAGCGCGTAA 1153
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
436 ProAspGlyLeuGlyAlaValProGlnGlnArgLeuProAlaGlyProSe 452
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1154 TCCGCTAGACATCTGCTACCTGCTTTGCGCGATTTAATCGTCGCG 1203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
452 rGlyArgArgHisProProPro.....ArgGlyArgSerArgA 466
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1204 GATACCGA.....CAGCGCGCAAGCAATTTGGTTG...CTTGA 1238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
466 rGlyAlaArgAlaCysArgAlaSerArgAlaSerValSerAlaGlyLeuAla 482
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1239 ATTGACGAAGAAGACCT 1256
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
483 AlaSerArgArgPro 488

```

seq\_name: sp\_virus:039782

seq\_documentation\_block:

```

ID O39782 PRELIMINARY; PRT; 867 AA.
AC O39782;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MEMBRANE GLYCOPROTEIN.
OS Equine herpesvirus 1.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10326;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BK343;
RA Kirisawa K., Kobayashi T., Kawakami Y., Iwai H.;
RT "Nucleotide sequences of open reading frames 1, 24 and 71 of an
RL J. Equine Sci 7:79-87(1996).
DR EMBL; D88734; BAA20038.1.
SQ SEQUENCE 867 AA; 86630 MW; C8DAD90D9B64407C CRC64;

alignment_scores:
    Quality: 106.50      Length: 460
    Ratio: 0.486        Gaps: 12
    Percent Similarity: 47.609      Percent Identity: 20.652

alignment_block:
US-09-303-518D-127/rev x O39782 ..

Align seg 1/1 to: O39782 from: 1 to: 867

1322 TCACGACCTTACGCAACAGCGGCCNNATTCGTATTGCCCCGGCAGAC 1273
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
50 SerSerSerProThrThrSerProThrThrSerSerSerProThr 66

1272 GAAGCTGCACAAAGCGAGGTCTTCTGTCCTCAATTCACCAACCAAC 1223
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
66 rSerThrHisThrSerSerSerSerThrThrGlnSerSerSert 83

1222 CTGCGCGGTGTCGGTATCG...CCGACGATTAAATCGCGCAAGCAGG 1176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
83 hrAlaAlaThrSerSerSerAlaProSerThrAlaSerSerThrThrSer 99

1175 GTAGGCAGGATGCTAGCGCATTCAGCGCTCGTAAGTACCAATCGGCAC 1126
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
100 IleProThrSerThrSerThrGluThrThrThrThrProThrAlaSe 116

1125 CATGCGCGGTGCGCCACCGTTCGCGTGTGCGTGAACCTTGAAGAGTTGT 1076
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
116 rThrThrThrProThrThrThrAlaAlaProThrThrAlaAlaThrT 133

1075 TTTTCAGGAATAGCGAGGTGCGTACGCGTGATGAGTATTGTCGCGC 1026
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
133 hrThrAlaValThrThrAlaAlaSerThrSerAlaGluThrThrAla 149

1025 TGGCGCGCAACCCAGCGCGAGCTCTTTCCTCGCGCCTTCTCGATAAC 976
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
150 ThrAlaThrAlaThrSerThrProThrThrThrThrProThrSerThr 166

975 GGAATCTGATTGTGCTAGCTCCCAATAATCATCGTCGCGCCTTGTGTA 926
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
166 rThrThr.....ThrAlaThrThrThrValP 175

925 TCGCGCGGTTCATACCGCAACCGGAATCAGCGGTGTCGCGCAAC 876
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
175 roThrThrAlaSerThrThrThrThrThrThrAlaAlaThrThrThr 191

875 AATTTCGCCGCGAGTAATTTCGATATCTTTCGACACCCCAACG.....GT 832
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
192 AlaAlaThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAl 208

831 ACGCAAGAGCGGTGTTTGTGACTTGAGAACCCCAACCAATTCAGC 782
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

208	aThrThrThraAlaAlaThrThrThraAlaAlaThrThrThraAlaAlaThrT	225
781	GCTCGGTTCAGACGGCTGTTGCAACAAACATCGCGATGGCAATTACA	732
225	hrThraAlaAlaThrThrThraAlaAlaThrThrThraAlaAlaThrThrT	241
731	TCTTGATTAATGATGGTCCAAACGGTTTGTTCGACCGCGCGCTCAAT	682
242	AlaAlaThrThrThraAlaAlaThrThrThraAlaAlaThrThrSerAl	258
681	GAATGAATGTCGTCGCCACCTCAACCGCGCGATCGCGCGCGCAATT	632
258	aThrThrAlaAlaThrThrThraAlaAlaThrThrThraAlaAlaThrT	275
631	CATGCTTCGATGTTGCGACAGATTTTCAGACGGC	597
275	hrAlaAlaThrThrThraAlaAlaThrThrThraAlaAlaThrThrT	291
596	ACGTCGCGCCAGCTGCCCTTACACACATCGATGGATTTACGGCTC	556
292	AlaThrThrThraAlaAlaThrThrThraAlaAlaThrThrThraAlaAlaT	308
555	GCTCAACGGCTCAATACACAGCANACNCTGCTGAATCNCNGCGGCTT	506
308	rThrThrAlaAlaThrThrThraAlaAlaThrThrThraAlaAlaThrT	325
505	CTTTGATCACACACAGCGTCTGCCGACGCGATGTTGTCATCGCA	456
325	hrAlaAlaThrThrThrGlySerProThrSerGlySerThrThrThrT	341
455	TTGACGACGATGGCAAGCGTTCGCATCGACGCGCA	409
342	GlyAlaSerThrThrThrProSerAlaSerThrAlaThrThrSerAlaThrPr	358
408	GAACGACGGNTACGACGCGAGTCCAAACCGGATTTGGATCAGATTGC	359
358	oThrSerThrThrSerAlaAla	367
358	NNCNNANTTCNTCGCGCTTAAGTTTGCAACGCTTCGGCGCGPAGGT	309
367	laThrThrSerThrProThrProThrSerAlaAlaThrSerAla	381
308	TCGAATCGATTCGTCGTGCTTCAACGGCAATCAGCAGCAGTGAAG	259
382	GluserThrThrGluAlaProThrSerThrProThrThrAsp	396
258	TAGCGCTTTTCGCCGCGATGATGCGCGGATTTGCTGAACNGCGG	209
396	rThrThrProSerGluAlaThrThrAlaAlaThrThrSerProGluSerThrT	413
208	CGGTAAC	192
413	hrValSerAlaSerThrThrSerAlaAlaThrThrThraAlaPheThrThrGlu	429
191	GGATNCTTTTGTCTCAACAGCACTTGGCTTTTTCAGCGCATCGCC	142
430	SerHisThrSerProAspSerThrThrGlySerThrThrAlaGluPr	446
141	TTCCCTTGACTTCATCAGGGCGCATACCGCATATTTCTTCGCCAAGCA	92
446	oSerSerThrPheThrLeu	460
91	ACGCGACTTCGGTAATGACGGCGCCCTCATAAATGACT	54
461	ThrAspGlnPheThrGlySerSerAlaSerThrThrGluSerAspSer	475
53	TCGTCGGTCTGCCCGCGATGGGC	30
476	ThrAspSerSerThrValProThrThrGly	485
seq_name: sp_invertebrate:076602		
seq_documentation_block:		

ID	076602	PRELIMINARY;	PRT;	1275 AA.
AC	076602;			
DT	01-NOV-1998 (TrEMBLrel. 08, Created)			
DT	01-NOV-1998 (TrEMBLrel. 08, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	HYPOTHETICAL 122.9 KDA PROTEIN.			
OS	H02F09.3.			
GN	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			
OC	Rhabditidae; Peloderinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RX	MEDLINE=94150718; PubMed=7906398;			
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,			
RA	Bonfield J., Burton J., Connell M., Cooper T., Coulson A.,			
RA	Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,			
RA	Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,			
RA	Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,			
RA	Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,			
RA	Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,			
RA	Smalidon N., Smith A., Sonhammer E., Staden R., Sulston J.,			
RA	Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,			
RA	Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.,			
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.			
RL	elegans.";			
RL	Nature 368:32-38(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RX	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.			
RA	Geisel C., Harmon G.;			
RL	"The sequence of C. elegans cosmid H02F09.";			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RA	Waterston R.;			
RL	Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF077538; AAC64622.1;			
KW	Hypothetical protein.			
SQ	SEQUENCE 1275 AA; 122924 MW; CB59B70C05959E25 CRC64;			
alignment_scores:				
Quality:		106.50	Length:	433
Ratio:		0.505	Gaps:	16
Percent Similarity:		48.730	Percent Identity:	20.092
alignment_block:				
US-09-303-518D-127/rev x 076602				
Align seg 1/1 to: 076602 from: 1 to: 1275				
1304	AGCGGCGCNTATTCGATTTCGCGCGGACGACGAGCTGCACAAAGCGAG	1255		
400	ThrAlaProSerThrValValThrValProThrThrValMetThrSerAr	416		
1254	GTCCTCT	1220		
416	gSerThrValThrThrProThrThrGlySerSerProSerThrA	433		
1219	CGCGGCTGCTGGTATCGCGACGATTAAATCGCCAAACAGCAGGTAGGC	1170		
433	laGlyThrSerLeuAlaSerThrAlaValThrGluThrSerIleGly	449		
1169	AGGATGCTAGCGCATTCACGCTCGTAAGTACCAATCGCACCATGGC	1120		
450	SerSerSerThrProLeu	455		
1119	CGCGTCGCCACCGTTCGCGCTGCTGAAGCTTGAAGAGTTTGTTCATCA	1070		
456	ProSerGlnSerThrSerLeuSerMetSerSerLeu	467		

1069 GGAATGGCCGAGGTCGTACGCTGATGAGTATTG.....TCC-1029  
468 .....SerThrTyThrProSerSerSer 475  
1028 GGCTGGGCGCAACCCAGCGC...AACAGCTCTTTGCTGCGCCCTTCCTC 982  
476 ThrAlaGlyAlaThrSerProAlaThrGlnSerThrLysProThrI 492  
981 GATAACGGAATCTGATTGTGGTAGCGTCCCAAAATAATGTCGCGCCTT 932  
492 eGlyThrSerMet.....SerSerGlyProT 501  
931 GTGTAATCGCGCGCTTCAATACCGAACCAGAAATCACGCGGTGCTCGC 882  
501 hrThrValAlaPro.....GlyAla 507  
881 TCACCAATTCGCCCGCAGTAATTTGCGACTTTTCGACCCCAAAACGGT 832  
508 SerThrGluSerThrValLeuGlnSerThrProSerGlyThrThVa 524  
831 ACACAAGAGCGTGTGTTTGTGACTTGAGACCCCAACCAATCAGCG 782  
524 lThrLeuProSerGlySerSerThr..... 532  
781 GCTCGGTGTCAGCGGCTGTGCAAAACAAACCTCCGATGGCAATTACA 732  
533 .....AlaThrAlaGlyThrSerProGlnAlaSerThr 543  
731 TCTTGATAATTGATGCTCCCAACGGTGTGTTGTCAGCGACCGCTCAAT 682  
544 .....ValThrThrValThrAspIleSerThrValSer.. 554  
681 GAAATGAATGTCGTGCCACTCAACCGCGCGGATGCGGCGCGCAATT 632  
555 .....GlySerT 557  
631 CATGTGTTTCGATGTCGACACATTTTCAGACGCGACGTCTGCGCCAGCT 582  
557 hrValThrSerGlnThrAlaGluSerSerLeuSerThrGluSerProThr 573  
581 GCTTTACACATGATGATTTTACGTCGCTCAACCGCTCAATACCAGCAN 532  
574 SerAlaGlySer.....SerIleSerThrValSerThrValSe 586  
531 ACNCTCGT.....CTGAATCNTGCGGCTTCTTTGATCACAA 494  
586 rSerGlnProSerThrTyIleProValSerSerAlaSerSerIleTyrS 603  
493 CCACAGGCTCTCGCGNAGCGGATTTGTCATCGCATTCGAGAGATG 444  
603 erThrLeuSerGlySerThrGlySerThrAlaSerProGlyThrThrGlu 619  
443 CGGAACGCTCGGATCGACGCGGAGGATTTGCTGAACGCGAGGNTACG 394  
620 SerSerGlySerSerThrSerGlyProSerThrIleSerGlySerSerAl 636  
393 CAGCGCATCCCAACACCGGATTCGATTCATGCTGCNNCANNNTTCGCG 344  
636 aSerThrVal.....ThrGlySerThrV 644  
343 CGCTTAAGTTTGCACGCTTCGGCGCGPAGCGTTTCGAACCTCGATTTCG 294  
644 alThrGluAlaSerThrIleSerGlySerThrGluSerSerThrIlePro 660  
293 TCGTTCCTTCAACGGCAATCAGACCGGCTGAAGTACGCGCTTTTCGCC 244  
661 Gly.....SerThrGluSerThrValSerGluAlaSerThrValSerG 675  
243 GCGATGATGCGCGGCTTTGCTGCTGAACGCGCGCGGTAACACACCGC 194  
675 ySerSerValSerThrValSerGlySerThr.....GluSerThrSera 690

193 CCGGATNCITTTTGTCTTCAACAGCAGCACTTGGCTTTTGTGACGCATCG 144  
690 laGlyAlaSerThrValSerGlySerThrGlySerThrValSerAspSer 706  
143 CTTTCCTGACTTTCATCAGGGGCGCATPACCGGCATATTTCTTCGCCAAG 94  
707 SerThrIleSerAspSerThrGly.....SerThrAsnAlaProG 720  
93 CAACGCGACTTCGTAATGACGGCGCGCTCATAAATGACTTCTCCTCGGT 45  
720 ySerThrGluSerThrValThrGlySerSerValSerThrValSerGly 736  
seq\_name: sp\_bacteriap:P73758  
seq\_documentation\_block:  
ID P73758 . PRELIMINARY; PRT; 574 AA.  
AC P73758;  
DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE ABC TRANSPORTER.  
GN SLR0864.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP  
RX MEDLINE=97061201; PubMed=8905231;  
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
MiyaJima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,  
Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Okumura S.,  
Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
Synechocystis sp. strain PCC6803. II. Sequence determination of the  
entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
DR EMBL: D90909; BAA17810.1; -  
DR InterPro: IPR003593; AAA.  
DR InterPro: IPR003439; ABC\_transportr.  
DR InterPro: IPR001687; ATP\_GTP\_A.  
DR Pfam: PF00005; ABC\_tran; 2.  
DR SMART: SM00382; AAA; 1.  
DR PROSITE: PS00211; ABC\_TRANSPORTER; UNKNOWN\_2.  
KW Complete proteome.  
SQ SEQUENCE 574 AA; 65027 MW; 2A75BC56BDA8FD17 CRC64;  
alignment\_scores:  
Quality: 105.50 Length: 300  
Ratio: 0.738 Gaps: 15  
Percent Similarity: 47.667 Percent Identity: 23.333  
alignment\_block:  
US-09-303-518D-127 x P73758 ..  
Align seg 1/1 to: P73758 from: 1 to: 574  
502 AAAGAAGCCGCGANGAT...TTCAGACGANGTNTCTGCTGTTTGTGAGCG 548  
270 GlnGlnAlaPheValAspLysPheArgAlaSerAlaThrArgSerThrG 286  
549 TTTTACCGAGCGTAAATCCATCTGTGTAAGGAGCTGGCGGAGAGTGC 598  
286 nAlaLysSerArgGluGlnLeuGluLysIleGluArgValGluAlaP 303  
599 CGTCTGAAATGCTGCCACATCGAAACACATGAATTCGCGCGCGCGCAT 648  
303 ro.....ValGlyGlyValArgThrLeuLysPheHisPhePro... 315  
649 CCGCGCGGTTTGTAGTGGCAGCGACATTCATTCATTGACCGCGTGGTGC 698  
316 ProAlaProArgSerGlyArgGlnValAlaLeuIleGluAspile..... 330







501 CAAGAAGCCGNCGGANGATTTCACAGCANGTNTGCTGTATTGACCGTT 550  
:::||||| ::||| |||  
478 ArgArgGLyArgSerArgSerArgThrProArgArgLy..... 490  
:::||||| ::||| |||  
551 TGACCGAGCGTAATAATCATCTGTGTAAAGCAGCTGGCGCAGACGTGCCG 600  
||| ||| |||  
491 .....ArgSerArgSerArgSerLeuV 498  
.....ArgSerArgSerArgSerLeuV 498  
601 TCTGAATGTGCACAACATCGAACAACATCAATTCAGATGTAAATTCGCATCGGCGT 650  
|| :: :||| ::||| ||||| |||  
498 alArgArgGLyArgSerHisSerArgThrProGlnArgArgGLyArgSer 514  
:::||||| ::||| |||  
651 GCGCGTGTGACTGGCACGCACATTCATTTCATTGAGCGCGTGCTGCAC 700  
||| ::| |||  
515 GlySerSer.....SerGluArgLysAs 522  
701 ACAAAACCCTTTGGACCATCAATTCAGATGTAAATTCGCATCGGCGT 750  
:::||||| ::||| ||||| |||  
522 nLysSerArgThrSerGlnArgSerArgSerAsnSerSerProGluM 539  
:::||||| ::||| |||  
751 TTCTTTGCAACAGCGCTCTGAACACCGAGCGCGTAATTCCTTTGGTGG 800  
:: :||| ||||| ::||| ::|||  
539 etLysSerArglleSerSerArgArgSerArgSer..... 551  
801 TTCTCAAGTCACAACAACACCGCTCTGCGTACCCTTTGGTGGCGAAG 850  
:::||||| ::|  
552 LeuSerSerProArgSerLysAlaLysSer..... 561  
851 TATCGCAAATTACTCGCGCGAATTGGTTGACGCAGACAAA..... 890  
562 .....ArgLeuSerLeuArgSerLeuSerGlyS 572  
891 .....CGCGTGATTTCCGCTTCGGTATTCAGACGGCGCATTCAC 929  
||| ::||| ::|||  
572 erSerProCysProLysGln.....LysSerGlnThrPro 583  
930 ACAAGCGCGCAGCATTAITGGGAGCTTACCACATCAGATTCCTCGTTA 979  
||||| ::||| ::|||  
584 ProArgArgSerArgSerGlySerGlnProLysAlaLysSerArgTh 600  
:::||||| ::|||  
980 TCGAAGAAGCCGCACCAAGAGCTTCCTGGTGGGTTCGCGCGCAGCCG 1029  
||| ::||| ::|||  
600 rProProArg.ArgSer.ArgSerSerSerProProLysGlnLy 616  
1030 GACAAATACCTCATCAGCTACGCCCTCGGCATTCCTCGAAAAACAA 1079  
:::||||| ::|||  
616 sSerLysThrProSerArg.....GlnSerH 625  
1080 ACTCTTCAAGTTACAGACGCCGTCAACGCTGGCGACCGCGCATCGTGC 1129  
:::||||| ::||| ::|||  
625 isSerSerSerProHisProLysValLysSerGlyThrProProArg 641  
11130 CGATTGTACTTACGACCGCGTAAATGCCGCTACACATCTGCTACCC... 1177  
:: :||| ::|||  
642 GlnGlySerlleThrSerProGlnAlaAsnGluInSerValThrProGl 658  
1178 .....TGCTTTGCGCATTTAATCTCGCGGATACCCGACAGCGC 1217  
||||| |||||  
658 nArgArgSerCysPhe.....GluSerSerProAspProGluLeu 672  
1218 GCAAGCATTTGGTTGCTTGGAAATGGACGAAGAAGACCTCGCTTTGTGCA 1267  
:::||||| ::|||  
672 vsSerArg.....ThrProSerArgHisSerCysSer 682

```
seq_name: sp_human:Q90HA8
seq_documentation_block:
ID Q90HA8 PRELIMINARY;
PRT; 2296 AA.
```

```

425 TCGA.....TGCCGAGCGGTTGCCATCTTC 450
    ||||
    :|||
654 gArgGlyArgSerArgSerArgThrProAlaArgGlyArgSerArgs 671
451 GTCAATCGGATGGACACCAATCGCTNGCGGACAGCCCTGTGGTGTGAT 500
    :|||
    :|||
671 erArgThrProAlaArgSerGlyArgSerArgSerArgThrProAla 687
501 CAAAGAGCGCGCGANGATTTCAGACGANGTNTGCTGGTATTGAGCGGTT 550
    :|||
    :|||
688 ArgArgGlyArgSerArgSerArgThrProArgArgGly..... 700
551 TGACCGAGCGCTAAATTCATGTGTGTAAGCAGCTGGCGCAGACGTCGCG 600
701 .....ArgSerArgSerArgSerArgSerLeuV 708
601 TCTGAAATGCTGCCACATCGAACAACATGAATTCGGCGCGCGCATCC 650
    :|||
    :|||
708 alArgArgGlyArgSerHisSerArgThrProGlnArgArgGlyArgSer 724
651 GGCCGGTTTGAGTGGCAGCAGCATTTTCATTGAGCGCGTGGTGCAC 700
    ||||
725 GlySerSer.....SerGluArgLysAs 732
701 ACAAAACCGTTGGACCATCAATTAATCAAGATGTAATGCCATCGGACGT 750
    :|||
    :|||
732 nLysSerArgThrSerGlnArgSerArgSerArgSerProGluM 749
751 TTTTGTGCAACAGCGCGTCTGAACCGAGCGGTGATTCCTTTGGGTGG 800
    :|||
    :|||
749 etLysSerArgIleSerArgArgSerArgSer..... 761
801 TTCTCAAGTCAACACACCGCTCTTGCTGCTACCGTTTGGTGGCAAG 850
    :|||
762 LeuSerSerProArgSerLysAlaLysSer..... 771
851 TATCCGCAATTAATCTGGCGGCGAATTTGGTGGACGACACAA..... 890
    :|||
    :|||
772 .....ArgLeuSerLeuArgArgSerLeuSerGlyS 782
891 .....CGCGTGATTTCCGGTTCGGTATTGAACGCGCGCATAC 929
    :|||
    :|||
782 erSerProCysProLysGln.....LysSerGlnThrPro 793
930 ACAAGCGCGCAGCATTTATTTGGAGCTTACCAATCAGATTTCCGTTA 979
    :|||
    :|||
794 ProArgArgSerArgSerGlySerGlnProLysAlaLysSerArgTh 810
980 TCGAAGAGCGCGCAGCAAGAGCTTTCGGCTGGTGGCGCGCAGCCG 1029
    :|||
    :|||
810 rProProArg.ArgSer.ArgSerSerSerProProLysGlnLy 826
1030 GACAATATCTCATCCAGCTACGACCTCGGCCATTTCTCGAAGAACAA 1079
    :|||
    :|||
826 sSerLysThrProSerArg.....GlnSerH 835
1080 ACTCTTAAGTTACGACGCGTCAACGGTGGCAGCGCGCCATGTGTC 1129
    :|||
    :|||
835 IsSerSerSerProHisProLysValLysSerGlyThrProProArg 851
1130 CGATTGGTACTACGAGCGGTATGCGCGTAGACATCTCCCTCCCTACCC... 1177
    :|||
    :|||
852 GlnGlySerIleThrSerProGlnAlaAsnGluGlnSerValThrProGl 868
1178 .....TGCTTTTTCGCGGATTTAATCTGTCGCGATACCGACGCG 1217
    :|||
    :|||
868 nArgArgSerCysPhe.....GluSerSerProAspProGluLeuL 882
1218 GCAAGCATTTGGTGTGGAAATGGACGAAGAACCTCGCTTTGTGCA 1267
    :|||
    :|||
882 ysSerArg.....ThrProSerArgHisSerCysSer 892
1268 GCTTCGTCT 1276

```

```

893 GlySerSer 895
    :|||
    :|||

```

```

seq_name: sp_bacteriap:Q99UM0

```

```

seq_documentation_block:

```

```

ID Q99UM0 PRELIMINARY; PRT; 435 AA.
AC Q99UM0;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE GLUCOSE-INHIBITED DIVISION PROTEIN GID.
GN GID OR SA1094 OR SAV1251.
OS Staphylococcus aureus (strain N315), and
OC Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158879, 158878;
RN [1]
SEQUENCE FROM N.A.
RP SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);
RC MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue K.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RA "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus";
RL Lancet. 357:1225-1240(2001).
CC -!- COFACTOR: FAD (BY SIMILARITY).
DR EMBL; AP003133; BAB42346.1; -.
DR EMBL; AP003361; BAB57413.1; -.
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR002218; GIDA.
DR InterPro; IPR000205; NAD_binding.
DR InterPro; IPR001100; pyr_redox.
DR PRINTS; PRO0368; FADPNR.
DR PRINTS; PRO0411; PNDRDTASEI.
DR ProDom; PD003738; GIDA; 1.
KW Complete proteome; FAD; Flavoprotein; Oxidoreductase.
SQ SEQUENCE 435 AA; 48371 MW; B76C254394051425 CRC64;

```

```

alignment_scores:

```

```

    Quality: 105.00      Length: 272
    Ratio: 0.784        Gaps: 13
    Percent Similarity: 49.265      Percent Identity: 21.324

```

```

alignment_block:

```

```

US-09-303-518D-127 x Q99UM0 ..

```

```

Align seg 1/1 to: Q99UM0 from: 1 to: 435

```

```

331 GCAAACTTAAGCGCGGANGANGAANTNGNGCAATCTGATCAATCCGGTTT 380
    |||||
    :|||
10 AlGlyLeuAlaGlySerGluAlaAlaTyrGlnLeuAlaGluArgGlyI 26
381 GTGAGCTGCGCTG...CGTANCGTCCGTTTCAGCAAAATCCCTGCCGCTG 427
    :|||
    :|||
26 eLysValAsnLeuIleGluMetArgProValLysGlnThrProAlaHisH 43
428 ATCCGAGCGGTTTCGCATCTTCGTC.....AATCGATGGACACCAAT 471
    :|||
    :|||
43 isThrAspLysPheAlaGluLeuValCysSerAsnSerLeuArgGlyAsn 59
472 CCGCTNGCGCAGACCCCTGTGGTGTGATCAAGAACGCGCGANGATT 521
    ||||
    :|||
60 AlaLeu...ThrAsnGlyValGlyValLeuLysGlu.....GluMe 72
522 CAGACGANGTNTGCTGTTGAGCGGTTTGACCGCGGTAAATCCATG 571

```

```

72   tArgArgLeuAsnSerIleIleLeuGluAlaAlaAsPlys..... 85
572  TGTGTAGGCAGCTGGCGCGCAGAGCTGCGCTGAAATGCTGCCAACATC 621
86   .....AlaArgValProAlaGlyAlaLeuAlaVal 96
622  GAAACACATGAATTCCGGCGCGCCGCATCCGCGCGGTTGTAGTGGCAGCA 671
97  AspArgHisAspPheSerGly..... 103
672  CATTCATTTCATTGAGCGCGTGCAGTGCAGCAAAACACGTTTGGACCATCA 721
104  ....TyrIleThrGluThrLeuLysAsnHisGluAsnIleThrValIleA 119
722  ATTATCAAGATGTAATGCCCATCGGA.....CGTTTGTGTTGCAACA 762
119  sn.....GluGluIleAsnAlaIleProAspGlyTyrThrIleIleAlaThr 134
763  GCGCTCTGACACCGGAG.....CGCTGATTGCTTTGGGTGG 800
135  GlyProLeuThrThrGluThrLeuAlaGlnGluIleValAspIleThrGI 151
801  TTCTCAAGTCAACAACACCGCCTCTTGCGTACCGTTTGGTGGCGAAG 850
151  lysasp.....GlnLeuTyrPheTyrAspAlaA 161
851  TATCGCAAAATTACTGGCGGGAATGGTTGACGAGACACACCGGTGATT 900
161  laAlaProIleIleGluLysGluSerIleAspMetAspLysValTyrLeu 177
901  .....TCGGTTCGTATTGTAACGCGCGGATTAC 929
178  LysSerArgTyrAspLysGlyGluAlaIleTyrLeuAsnCysProMetTh 194
930  ACAAGCGCGCAGCATTAATTGGGACGCTACCAACATCAGATT..... 972
194  rGlu.....AspGluPheAsnArgPheTyrAspAlaValLeuGluA 208
972  ..... 972
208  laGluValAlaProValAsnSerPheGluLysGluLysTyrPheGluGly 224
973  .....TCGGTTATCGAAGAGCGCGCAGCAAGAGCTGTT 1007
225  CysMetProPheGluValMetAlaGluArgGlyArgLysThrLeuLeuPh 241
1008  CGGCTGGTTCGCGCG 1023
241  eGlyProMetLysPro 246
seq_name: sp_mammal:O18758
seq_documentation_block:
O18758: PRELIMINARY; PRT; 13288 AA.
AC O18758:
101-JAN-1998 (TrEMBLrel. 05, Created)
01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
SUBMAXILLARY APOMUCIN.
Sus scrofa domestica (domestic pig).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleoc
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
NCBI_TaxID=9825;
[1]
SEQUENCE FROM N.A.
MEDLINE=98070526; PubMed=9407109;
Eckhardt A.E., Timpte C.S., DeLuca A.W., Hill R.L.;
"The complete cDNA sequence and structural polymorphism of
polypeptide chain of porcine submaxillary mucin.";
J. Biol. Chem. 272:33204-33210(1997).
EMBL; AF005273; AAC62527.1; -.
DR InterPro; IPR000359; Cys_Knot.

```

```

770 AGACGGCTCTTGC.....AAACAAACGTCGATGGCAATTAACATCTTG 727
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12546 ysThrGluAlaThrThrPheProGlyGluAsnGluThrThrArgValGly 12562
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
726 ATAATTGATGGTCAACACGGTTTGTTCACCGACCGGCTC.....AA 683
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12563 IleAlaThrGlyThrThrGlyIleValSerArgLysThrGluProG1 12579
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
682 TGAATGAATGTCGGTGCACCTCAACACC...GCCGGATCGCGCGCG 636
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12579 ySerTyrAsnThrGluAlaThrThrSerIleGlyArg.....12591
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
635 AATTCATGCTTTCGATGTGGACGATTTTCAGACGACGCTCTCGCC 586
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12591 .....12591
585 AGCTGCTTACACATGATGATTTAGCTCGGTCAACG.....GC 545
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12592 SerGlyThrThrHisThrAspLeuProGlyGlyThrIleValLeuPr 12608
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
544 TCAATACACAGCANACNCTGCTGAAATCNCGCGCTCTTTGATCACA 495
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12608 oglyPheSerHisSerGln.SerSerLysProGlySerSerValThr 12624
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
494 ACCACAGGCTGCGCCNAGCGGATGGTGTCCATCGCA.....456
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12625 ThrProGlySerProGluSerGlySerGluThrGlyThrSerGlyGluPh. 12641
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
455 .TTGACGAGATGGCGACGCGTCCGATCGACGACGAGGATTTGCTGA 407
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12641 eserThrThrValIleSerGlySerSerHisThrGluAlaThrThrPheI 12658
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
406 ACGGAGGNTACGACGCGAGTCCCAACACCGGATGGATGATGCGNN 357
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12658 leGlyGlySerGlySerPro.....Gly 12665
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
356 CNNANTTCNCGCGCTTAAGTTTCCCAACGCTTCGGCGCGTAGCGTTC 307
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12666 ThrGlySerArgProGlyThrThrGlyGluLeuSerGly.....Th 12679
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
306 GAATCGATTTGCTGCTTCAACGCGCATCATCAGCACCGCTCAAGTA 257
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12679 rThrIleAlaSerGlyAsnAlaThrThrGluAlaThrThrSerThrGluT 12696
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
256 CGCGTTTTCGCGCGATGGATGGCGGATTTTCCCTGAACNGCGCG 207
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12696 hrArg.....IleGlyProGluThrGlyAla 12704
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
206 GTAAACACACGCGCGATNCTTTTGTCTTCAACACGACTTGGCCTTT 157
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12705 GlnThrThrValProGlySerSerGlnValSerGlySerGluThrGlyThrSe 12721
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
156 TTTGACGCGATCGCTTCTCTTACTTTTCATCAGGCGGCGATACCGCAT 107
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12721 rGluAlaValSerAsn.....ProAlaI 12729
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
106 ATTCTTCCCAAGCAACCGCTTCGGTAATGACGGCGCGCTCA 63
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12729 leAlaSerGlySerSerThrGlyThrThrSerGlyAlaSer 12743
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
seq_name: sp_invertebrate:Q9NK94
seq_documentation_block:
ID Q9NK94 PRELIMINARY; PRT; 581 AA.
AC Q9NK94;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 64.2 KDA PROTEIN.
GN BG:DS07295.4.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```

```

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, CN BW SP;
RX MEDLINE=99403001; PubMed=10471707;
RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
RA Celniker S., Rubin G.M.;
RT "An exploration of the sequence of a 2.9-Mb region of the genome of
RT Drosophila melanogaster: the Adh region.";
RL Genetics 153:179-219(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, CN BW SP;
RA Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA Nethi K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA Szieran L., Snir E., Svirskaas R.R., Wan K.H., Weinburg T., Zhang R.,
RA Zieran L.L., Rubin G.M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; A5003412; AAF44927.1; -.
DR FlyBase; FBgn0028861; BG:DS07295.4.
KW Hypothetical protein.
SQ SEQUENCE 581 AA; 64222 MW; 44841PDI6526701 CRC64;

```

```

alignment_scores:
  Quality: 104.50      Length: 409
  Ratio: 0.601        Gaps: 22
  Percent Similarity: 42.543      Percent Identity: 23.472

```

alignment\_block:

US-09-303-518D-127 x Q9NK94 ..

Align seg 1/1 to: Q9NK94 from: 1 to: 581

```

66 CGGCGCGCTCATACCAAGTCCGCTGCTGGCGA.....AG 103
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
212 ArgLeuGluHisTyrArgAsnHisLeuArgTrpHisValThrProSerAr 228
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
104 AATATGCCGCTATCGCCCTN.....125
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
228 gLeuGluAsnPheProLeuGlyLeuAlaAsnThrAsnThrSerLysL 245
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
126 .....GATGAAGTCAAGGAGGCGTCCGCTCAAAAAGGCCAAGT 167
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
245 euLeuValAspAlaAlaPheTrpLeuLeuLeuThrGlnLysGluLysTyr 261
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
168 GCTGTTTGAAGCAAAAAGNATCCGGCGTGGTGTTCACGCGCCNGTTT 217
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
262 LysLysLysAsnGlnIleArgSerGluLysAsnThrLeuLysIleGluAr 278
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
218 CAGGCAA.....AATCGCGCGCATCCATCGCGCGGCGAAAGCGCGTACTT 261
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
278 gGluGluIleLeuAsnLeuArgLysLysLeuSerLeuLysCysGlyGluS 295
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
262 CAGTCGCTGCTGATTCGCGTTGAAGGCAA.....CGACGAATCCA 302
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
295 erCysGlyAsnAsnTrpArgIleArgLysGlyMetProSerArgAsnArg 311
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
303 GTTCGAACCTACGCGCCCGCAAGCGTGTGCAAACTTAAGCGCGGANGAAN 352
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
312 AlaValGlyLeuGlySerGlyGlyIleArgSer.....322
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
353 TNNNGNCAATCTGATCCCAATCGGTTTCTGGACTGCGCTGCTGCTGCTGCT 402
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

323 .....ValHisHisValAspProPheGlyAlaSerC 333
403 CCGTTTCAGCAAAATCCCTGCGCTGCATGCGCGAGCGCTTCGCCATCTTCGT 452
333 ysProGlyGlnValAspCysGlyAsnCys..... 342
453 CAATGCGATGGACCAATCCGTGTCGTCGCGGAGACCCCTGTGTTGTGATCA 502
343 .....AsnTrpLysThrAlaGlnGlyLys...ThrGlyGluAsnPh 355
503 AAGAAGCCGNCGANGATTTCAGACGANGTNTGCTGGTATTTCAGCGCGTTT 552
355 eAlaAlaAArgMetGlyGlnThrGlyGlnAlaGlyCysAspGlnHisA 372
553 ACCGAGCGTAAATCCATGTGTGTAAAGCAGCACTGGCGCAGAC..... 594
372 spLeu.Arg...IleHisLeuValValAlaAlaAspGluSerSerAspG1 387
595 .GTGCCGCTCTGAAAATCGTCCACATCGAAACACATGAATTCGGCGGCC 643
387 yGlnProAsnGluAspAlaArgAspAspProAspHisGluAspGlyGlyH 404
644 CGCATCCGGCGCGTTCAGTGTGCACGCACATTCATTTTCATTCAGCGCGTC 693
404 iSGlyAlaAsnAsnLeuSer.....Profile 412
694 GTGTCA.....AACAAACCGTTTCGACCACATCAATTA 725
413 ProAlaGluGlyHisProValLysAspLysSerGluTrpSerLeu.... 427
726 TCAAGATGTAATTCGCATCGGACGCTGTTGTCGAACAGCGCGCTCTGAACA 775
428 .....IleAsnValGlyAlaValPhe.....TyrAsnV 437
776 CCGAGCCGCTGATTGCCTTGGTGGTCTCTCAA..... 807
437 alPhePheValPheTrpLeuGlyGlnSerGlnHisLeuAsnGlyArgGly 453
808 ..... 809
454 GlySerAspGlnValGluGlyArgSerIleLysAlaGlnProArgProgl 470
810 CAACAACACCGCCTCTTCGTACGCTTTGGTGGCGAAGTAGTATCGCAA. 858
470 uAsnLysProSerHisIleArgThrSerGlnGluGlnAspGlnAsnGlnA 487
859 ..ATTACTCGCGGGCAATTGGTTGACGACGACAAACCGCGTGATTTCCGGT 906
487 spGlnProAlaCysGlnLeuGlyAsnGlySer.....GlnSer 499
907 TCGGTATTGAACGGCGCATACACAAGCGCGCACAGTATTATTGGGACG 956
500 AlalysMetGlyGlyAlaValGlyGluGlyAla..ThrLysIleTrp.... 514
957 CTACCACAATCAGATTCGTTTATCGAAGACGGCGCACCAAGACTGT 1006
515 .....ArgLysAlaAla..... 518
1007 TCGGCTGGGTTTCGCGCGCAGCGGCAATACTTCATCACGGGTACGACC 1056
519 .....AsnArgThrGluThrLysThrSerGlnArgThrArgPr 531
1057 CTCGGCCCATTCCTGAAACAAACAACTCTTCAGTTCCACGACAGCCGTCAA 1106
531 oLys.....GluArgGlyIleGlnArgGlnProGluT 542
1107 CGGTGGCGACCGCGCATGGTGC 1129
542 hr...GluThrAlaAlaThrCys 548

seq_name: sp_human:043419
seq documentation_block:

```

```
240 GluThrSerIlePheValSerMetThrSerAlaThrThrProSerGlyAr 256
864 AGTAATTTGGCATACTTTTCGACCCAAACGGTACGCAAGAGCGTGGTT 815
256 g.....ProThrPheThrSerThrValAsnThrProThrArgSerL 270
814 TGTGACTTGAGAACCA.....798
270 euLeuThrSerPheProThrHisLeuPheSerSerMetSerGlu 286
797 .....CCCAAGCAAT 787
287 SerSerAlaGlyThrThrHisThrGluSerIleSerSerProAlaLath 303
786 CACGCGCTCGGTTCAGACGGCTGTGCACAAACGTCGATCGCA 737
303 rThrSerThrLeu.....HisThrThrAlaGluSerProSerCyst 318
736 TTACATCTTGATAATTGATGGTCCAAACG.....GTT 705
318 hrThrThrThrSerPheIleThrSerThrThrMetGluProLeuSerThr 334
704 TTGTTTCCACCGACCGCCTCAATGAATGAATGTGCGTCCACTCAACC 655
335 ilValAlaThrThrGlyThrValLys.....343
654 GGCGGATGGCGCGCGCAATTCATGTGTCGATGTCGACGATTTT 605
344 .....ThrThrValThrSerSerThrAlaThrPheA 354
604 CACAGCGCAGTGTGCGCGCGCTTACACATGATGATTTTACGCTCG 555
354 rgGluThrThrLeuThrSerThrThrAspLeu.....SerThrGluSer 369
554 GTCAACAGCGCTCAATACACGACGACGCTGTAATCTGTCGCGCTTC 505
370 LeuMetThrAlaMetThrThrArgLeuThrSerAlaIleThrSe 386
504 TTGTATCAACAACACAGGCTCTGCCGACGATGTCGATCGCAT 455
386 rLysThrThrLeuThr.....SerLeuLysThrThrAlaL 398
454 TGACGAAGATGGCAGCGCTGGGATCG.....ACGCGAGGATTTTG 411
398 exArgProThrAlaAsnSerThrLeuSerSerLeuThrSerSerIleLeu 414
410 CTGACGCGAGGTAGTACGACGAGTCCACAAACGCGATGATCAGATT 361
415 .....SerThrThrLeuValProSerThrAspMetIleThrSe 427
360 GCNCCNANT.....TCNCGCGCTTAAGTTTGCCACGCTT 323
427 rHisThrThrAsnLeuThrArgSerSerProLeuLeuAlaThrLeuProT 444
322 CGGCGCGTGGTTCGACGCTGATTTCG.....TCG 291
444 hrThrIleThrArgSerThrProThrSerGluThrThrTyProThrSer 460
290 TTGCTTCAACGGCAATCACACGCGCTGAAGTACGCGCTTTTCGCGCG 241
461 ProThrSerThrValLysGlySerThrThrSerIleArgTyThrSe 477
240 ATGGATGGCGGATTTTGCCTGAACGCGCGGTAAACACACGCGCG 191
477 rMetThrGlyThrLeuSerMetGluThrSerLeuProProThrSerSers 494
190 GATNCTTTTGTCTCAACAGCACTTGGCCTTTTTCACGCGATCGCT 141
494 erLeuProThrThrGluThrAlaThr.....ThrPro 504
140 TCCTTACTTCATNAGGCGCATCCGCGATATTTCTCGCAAGCAA 91
505 ThrThrAsnLeuValThrThrGluIleThrSerHisSerThrProSerPh 521
```

```
90 CGCGACT.....TCGTAATGACGGCCCGTCATAAATGACTT 53
521 eserSerThrIleHisSerThrValSerThrProThrThrValValS 538
52 GTCGCGGTCTGCCGCGATGGC 30
538 erSerGlyProProThrSerGly 545
seq_name: sp_human:Q9H195
seq_documentation_block:
ID Q9H195 PRELIMINARY; PRT; 901 AA.
AC Q9H195;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE MUC3B MUCIN (FRAGMENT).
GN MUC3B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pratt W.S.; Crawley S.; Hicks J.; Ho J.; Nash M.; Kim Y.S.; Gum J.R.;
RA Swallow D.M.;
RT "Multiple transcripts of MUC3: Evidence for two genes, MUC3A and
RT MUC3B."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ291390; CAC19572.1; -.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000082; SEA.
DR SMART: SM00181; EGF; 2.
DR SMART: SM00200; SEA; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS50024; SEA; 1.
FT NON_TER 1
SQ SEQUENCE 901 AA; 96192 MW; F98FC817494ECD99 CRC64;
```

```
alignment_scores:
Quality: 104.50 Length: 444
Ratio: 0.522 Gaps: 14
Percent Similarity: 45.045 Percent Identity: 21.171
alignment_block:
US-09-303-518D-127/rev x Q9H195 ..
Align seg 1/1 to: Q9H195 from: 1 to: 901
1340 CTTTCCTTCGCAANGTTTCCAGCACCTTAGCAACAGCGGCNTATTC 1291
||||| |||
1 ProSerPheMetSer.....Se 6
1290 GTATTTCCCGGCGACAGCAAGTGCACAAAGCAGGCTCTTCGTCCTCA 1241
| : : : : ||| : : : : ||| : : : : ||| : : : :
6 rIleIleThrThrGluThrThrSerHisSerThrProSerPheSerSers 23
1240 ATTCCAAGCAACCAATGCTTGGCGGTGTGCGTATCGCGACGATTA 1191
: : : : : : : : : : : : : : : : : : : : : :
23 erThrIleHisSer.....ThrValSerSerSerThrThr 34
1190 TCGCGCAAAAGCAGGTAGCAGCATGTAGCGGATACGCGCTCGTA 1141
: : : : : ||| : : : : : : : : : : : : : : :
35 AlaIleThrSerProPheThrThrAlaGluThrGlyValThrSerThrPr 51
1140 AGTACCAATCGGCACCATGGCGGGTGCACCGTTGACGGCTGTCGTGA 1091
: : : : : : : : : : : : : : : : : : : : : :
51 oSerSerProSerSerLeuSerThrAspIleProThrThrSer..... 65
1090 ACTTGAAGAGTTTGTTCAGGAAATGCGCGAGGTCGTACGCGGTGATG 1041
: : : : : : : : : : : : : : : : : : : : : :
```

Mon Jul 1 09:25:31 2002

```
66 .....LeuArgThrLeu 69
1040 GAGTATTGTCGGCTGGCGGACCGACCGAGCTCTTGTGCTGCG 991
      ||||| :|||: :|||: :|||: :|||: :|||: :|||: :|||:
70 ThrProLeuSerLeuSerThrSerThrSerLeuThrThrThrAspLe 86
990 GCCTTCTTCGATACAGGAATCTGATGTGTAGCGTCCCAAAATACGT 941
      ||||| :|||: :|||: :|||: :|||: :|||: :|||: :|||:
86 uProSerIleProThrAspIle.....Sers 95
940 GCGCGCTTGTGTAATCGCGGCTTCATACCGACCGGAATCAGCGG 891
      ||| :|||: :|||: :|||: :|||: :|||: :|||: :|||:
95 erLeuProThrProIleIleSerSerProIleSerIleGlnSer 111
890 TTGCTGCGCTCA.....ACCAATCGCCCGCAGTAATTTGCGA 853
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
112 ThrGluThrSerSerLeuValGlyThrThrSerProThrMet..... 125
852 TACTTTCGACCCAAACAGGTACGCAAGGCGTGTGTGTGACTTGAG 803
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
126 .....SerThrValArg..... 129
802 ACCACCCAAAGCAATCAGCGCTCGTTCAGACGGCTGTTGCCAAAC 753
      ||| :|||: :|||: :|||: :|||: :|||: :|||: :|||:
130 .....AlaThrLeuArgSerThrGluAsnThrProIleSerSer 142
752 AAACGTCGATGCAATTACATCTTGATAATTGATGTCACAAACGGTTT 703
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
143 .....PheSerThrSerIle 147
702 GTTTGCACGACGGCTCAATGAATGATGTGCGTGCACACCAACGG 653
      : : : : : : : : : : : : : : : : : : : : : :
147 evalValThrProGluThrProThrThrGlnAlaProValLeuMets 164
652 CCGGATCGCGCGCGCGAATTCATGTTCGATGTGGCAGCATTTCA 603
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
164 erAlaThrGlyThrGlnThrSerProValProThrThrValThrPheGly 180
602 GACGGCAGCTCTCGCGCAGCTGCTTACACACATGGATTTTACGTCGGT 553
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
181 SerMetAspSerSerThrSerThrLeuHisThrLeuThrProSerThrAl 197
552 C.....AAAGGCTCAATACAGACANACNTCGCTGGAATCNTCGNCGG 509
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
197 aLeuSerLysIleMetSerThrSerGlnPheProIleProSerThrHisS 214
508 CTCTTTTGATCACACACAGGCTCTGCCGNAGCGGATTTGTCATC 459
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
214 erSerThrLeuGlnThrThrProSerIleProSerLeuGlnThrSerLeu 230
458 GCATTGACG.....AAGATGGCGAAGCGTCGGCATC 427
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
231 ThrSerThrSerGluPheThrThrGluSerPheThrArgGlySerThrSe 247
426 GACGGCAGGATTTGCTGAACGACGGNTACCGCAGCGATCCACAAAC 377
      |||| :|||: :|||: :|||: :|||: :|||: :|||: :|||:
247 rThrAsnAlaIleLeuThrSer..... 254
376 CGGATTGGATCAGATGCNNNNTTCNTGCGCTTAAGTTTGCCAAAC 327
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
255 .....PheSerThrIleIleThrProSerThrProThrIleIleMetSer 269
326 GCTTCGGCGCGTAGCGTTCGAACTCGATTTGCTGCTTCGCTTCAACGGC 277
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
270 Ser.....SerProSerSerAlaSerIleThrProValPh 281
276 AATCAGCAGCAGTGAAGTACGCGCTTTTCGCCGCGATGATGGCGCGA 227
      ||||| :|||: :|||: :|||: :|||: :|||: :|||: :|||:
281 eAlaThrThrIleHisSerValProSerSerProTyrIlePheSerThr. 297
226 TTTTGCCTGAAACNGCGCGGTAAACACACCGCGGATNCTTTTGTCT 177
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
298 .....GluAsnValGlySerAlaSerIleThrAlaPheProSerLeuSer 312
```